MODES OF GOVERNANCE AND PUBLIC SERVICE EFFICIENCY

by

CARLOS MIGUEL DE CASTRO SILVEIRA COELHO

A thesis submitted to
The University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

Institute of Local Government Studies
School of Public Policy
The University of Birmingham
May 2007

Abstract

This thesis examines the theoretical and empirical relationship between modes of governance and public service efficiency. We argue that different modes of governance yield different levels of efficiency depending on the nature and scale of the transactions upon which they are deployed. The experience of OECD countries is used to examine the effects of different modes of governance on the efficiency of education, health, and social protection systems.

In the education sector, the share of public providers is found to exert a negative effect on efficiency whereas the degree of decentralisation of the decision-making procedures of public providers is found to exert a positive effect on efficiency.

In the health sector, the introduction of market-type mechanisms to public integrated health systems is shown to have positive effects on efficiency, whereas further movement towards a market model of health care insurance and provision is shown to depress efficiency.

In the social protection sector, we conclude that as public social security systems exceed their remit to assist individuals smooth their income across the life cycle and/or states of nature and to provide basic social safety nets to the destitute, the efficiency of social transfers in reducing poverty is damaged.

This thesis was written in memory of my grandfather Fernando Martins Coelho

Acknowledgements

I would like to express my profound gratitude to both my supervisors, Peter Watt and Chris Skelcher, for their invaluable contribution to this work.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

This thesis examines the theoretical and empirical relationship between modes of governance and public service efficiency. There has been much debate in the political science and public management literature on the governance of public services, but relatively little conceptual or empirical work on the relationship between governance and efficiency. The economics literature, in contrast, has developed complex theoretical frameworks for understanding the nature of this relationship, and sophisticated methodologies for measuring efficiency. However, it still lacks a unified theoretical framework, and progress is yet to be made in investigating the empirical link between public service governance and efficiency. This thesis makes a contribution by bridging this gap. First, it brings together new institutional economic (NIE) theories of organisation and resource-based/evolutionary perspectives of the firm to clarify the nature of the relationship between governance and efficiency. Second, it draws on the experience of OECD countries operating similar services under different organisational configurations as a laboratory of inquiry to establish the impact of different modes of governance on public service efficiency.

The importance of public service efficiency cannot be overstated. Public service efficiency affects economic performance through two different channels: (i) via productivity changes in the public sector itself, and (ii) via effects on the private sector. Public services are responsible for a large proportion of the overall output of an economy. Public service productivity (and efficiency[1]) is therefore an important driver of average productivity nationwide. Furthermore, the organisation and functioning of public services also affects productivity in the private sector. On the one hand, taxes required to fund public services distort relative prices in the economy influencing economic incentives, namely the willingness to work and invest. On the other hand, goods and services such as education, health, and social protection are known to have a positive influence on the development of economic activity[2]. The extent to which public expenditure has positive growth effects through this second channel depends not only on the size of public expenditure but also on the efficiency of public expenditure.

Since the beginning of the 1980s, there has been a growing concern amongst most developed nations with the potentially negative consequences of a rising trend of government expenditure and taxation. This concern has prompted widespread

[1] One of the components of productivity.
[2] See Afonso, Ebert, Schuknecht and Thone (2005) for a good review of this literature.

adoption of various initiatives aimed at curbing the enlargement of the public sector, or at least bringing its growth to a halt (Joumard, 2002). The need to scale down the size of government and increase the efficiency of government activity has also raised an important question: How can policy influence efficiency? what is the relationship between governance and efficiency? This question is the key driver of the work presented in this thesis.

Our main theoretical assertion, rooted in a combination of new institutional economic theories of organisation and resource-based/evolutionary perspectives of the firm, is that different modes of governance yield different levels of efficiency depending on the nature and scale of the transactions upon which they are deployed. Thus the impact of different governance/organisational arrangements (including different degrees of public sector interference) on efficiency depends on the specific characteristics of the service/activity under consideration.

Our subsidiary research hypothesis is that improving the efficiency of government activity implies that (i) governments retreat from interfering with transactions that do not exhibit the kind of attributes that call for public central coordination; (ii) when public central coordination is required (e.g. education, health, and social security), it should be confined to the specific attributes of transactions that markets fail to address (public good attributes, merit good attributes, externalities, and/or information asymmetries), and in a way that explicitly acknowledges the risks of government failure. This implies (amongst other things) that once a given optimal structure of provision is defined, the units (public or private) that deliver the service/activity should be managerially decentralised/autonomous.

The first part of the thesis (chapters 2, 3, and 4) examines the relationship between governance and efficiency. The political science and economic literature on governance and efficiency is reviewed, the framework that underpins our main theoretical assertion is presented, and the research hypothesis is formulated.

The second part of the thesis (chapters 5, 6, and 7) draws on the international public sector reform experience to test empirically the research hypothesis in the context of three key areas of the welfare state - education, health, and social protection.

Chapter 2 reviews the most significant contributions of political science and economics to understanding governance and its relationship with efficiency.

Governance is an eclectic term deployed in various contexts with diverse and often contradictory meanings. In the political science literature the terms *governance* and *modes of governance* have been broadly used to refer to the steering and coordination of interdependent actors based on institutionalised rule systems (Benz, 2004). In the economics literature, modes of governance are seen as forms through which order is achieved and sustained in economic and social life. The most traditional and well established economic theories of organisation are rooted in NIE (including transaction costs economics and property rights theory). Other competing explanations of organisation are resource-based/evolutionary perspectives of the firm. Serious divergences divide these alternative explanations, and an integrated economic theory of organisation is still missing.

In the most conventional approach to economic organisation, governance is conceptualized as "the means by which to infuse order, thereby to mitigate contractual hazards and realize mutual gains from trade…[i.e.] the deployment of alternative modes of governance among differing transactions with the purpose of economizing on transaction costs" (Williamson, 2004, p.0). Modes of governance in turn are understood as institutional arrangements within which a transaction or set of related transactions are decided upon and implemented (Ménard, 2005)[3]. In this context, firms are seen as an instance of market failure - legal, contractual entities, owners of physical assets providing investment incentives and facilitating the resolution of hold-up problems.

The capabilities perspective, resource or competence based view of the firm is an alternative conception of economic organisation that sees the problem of economic coordination as an attempt to find integration mechanisms that sustain the division of labour amongst agents with incomplete, dispersed and disparate knowledge and help create and test new knowledge (Piore, 1992; Loasby, 1998a, 1998b; Kogut, 2000). Firms are conceived of as a bundle of essentially tacit "capabilities", shared by the human assets that comprise it. The boundaries of the firm are "partly determined by the differential between the capabilities they hold in relation to potential supplies and the relative costs of developing capabilities internally against accessing them via non-hierarchical means" (Araujo *et al.*, 2003, p.1263). Opportunism and threats of hold-up

[3] Recently Williamson defined a governance structure as "the institutional matrix in which the integrity of the transaction is decided" (Williamson, 1996, p.378).

are part of the calculus of deciding which is the most cost-effective way to access capabilities (Barney, 1999).

NIE deviates from the standard neoclassical assumptions regarding perfect information, unbounded rationality, and costless/instantaneous transactions. Incomplete information and limited mental capacity makes individuals face uncertainty about future events and incur transaction costs. To reduce risk and transaction costs individuals create institutions and develop modes of organisation embedded in those institutions (Ménard and Shirley, 2005). Markets, firms and hybrids are the archetypes of organisation, and the second half of the chapter is devoted to reviewing individually the NIE literature on these three modes of governance.

The study of the internal characteristics of hierarchies and their costs by new institutional economists remains an underdeveloped area. Whether or not the transaction costs apparatus is helpful for understanding the fundamental characteristics of the different modes of governance that determine their comparative efficiency in coordinating different kinds of transactions is the subject of controversy. Some authors argue that transaction costs concern exclusively market exchanges, and that the analysis of administrative/bureaucratic costs requires employing other research tools (Demsetz, 1988a; 2002). Other authors lay emphasis on the need to better identify the costs involved in the trade-off among the different organisational arrangements (Masten *et al.*, 1991; Joskow, 2005).

Following Hayek's seminal distinction between *spontaneous self generating order* and *organized ("made") order* we argue that economic organisation is essentially about bringing order into systems of human interaction. Order can be accomplished through a multitude of modes of governance, ranging from conscious deliberate design to spontaneous self-reinforcing rules. Each of these modes of governance entails costs of coordination that vary with the attributes of the transactions upon which they are deployed. Minimising these costs is the main problem of economic organisation.

Chapter 3 presents the theoretical framework and research hypotheses on the relationship between governance and efficiency that are empirically tested in the second part of the thesis.

Economic organisation is about bringing order into systems of human interaction. Order can be accomplished through a multitude of modes of governance, a continuum ranging from markets to hierarchies, i.e. from conscious deliberate design to spontaneous self-reinforcing rules aggregating elementary actions guided only by individual criteria of interest.

Achieving efficiency in the coordination and motivation of the decisions and actions of producers and consumers depends on how the information required for optimal allocation of resources and subsequent adaptations is managed. In theory two general solutions are feasible. Either transmit the dispersed information to a central planner who is then left with the task of computing the data and providing a solution to the resource-allocation problem, or else put in place a more decentralised system that involves less information transmission, delegating at least part of the decisional power to those who actually hold local information.

Different modes of governance yield different levels of efficiency depending on the nature and scale of the transactions upon which they are deployed. In other words, different modes of governance entail different costs in coordinating transactions, and these costs should be the decisive criterion to determine which transactions ought to be coordinated through markets and which should instead be centrally coordinated.

On the one hand, no system can solve the problem of coordinating economic activity at the societal level more efficiently than a system of markets (Arrow and Debreu, 1954). The price mechanism requires the transmission of less information than any other system capable of producing an equally effective outcome. Individuals and organisations acting on a self-interested basis, combining their local knowledge with the information synthesised by prices, and attempting to maximise their utilities and profits are motivated to carry out a plan that will lead them to an efficient outcome.

On the other hand, decentralised coordination of transactions with *design* and/or *innovation attributes* (Milgrom and Roberts, 1992) is sub-optimal due to the brittleness of the system in dealing with imperfect information and excessive demands in terms of communication. The quest for efficiency leads these types of transactions to be coordinated through formal organisations structured along hierarchical lines. Decentralised coordination of transactions with public good attributes, externalities, merit good attributes, and/or information asymmetries is also sub-optimal and leads

perfectly competitive markets to produce inefficient outcomes. Under these circumstances, *public* central coordination is required to deal with these particular features of transactions and to generate an efficient allocation of resources.

The case for public central coordination needs to be balanced with the idiosyncratic limitations of public bureaucracies and the political process. An imperfect market must be compared with an imperfect government. The relevant choice is not so much between an ideal norm (pure and perfect competition) and an existing imperfect institutional arrangement (the "real" market), but rather between alternative real and imperfect institutional arrangements (the market with or without external corrections). Numerous characteristics specific to political bureaucracies hinder the efficiency of governmental coordination. These need to be taken into account when calls for public central coordination are being made (Stiglitz, 2000).

Furthermore, in seeking to improve efficiency, individuals have kept forging innovative hybrid governance arrangements resting on intermediate levels of (de)centralisation of information, in between the two extremes defined by fully decentralised organisational systems entirely driven by market forces on the one hand, and fully centralised hierarchical/bureaucratic authority-driven organisations on the other. Alliance/network modes of governance are one specific example of such hybrid arrangements.

The structure and efficiency of public expenditure mediate the relationship between government size and economic performance. Apart from being limited to a set of core activities, and confined to certain levels, public expenditure needs to be efficiently administered if it is to have a positive effect on economic growth. Based on the theoretical framework exposed in chapter 3, we hypothesise that maximizing efficiency implies that (i) governments retreat from interfering with transactions that do not exhibit the kind of attributes that call for public central coordination; and (ii) when public central coordination is required (e.g. education, health, and social security), it should be confined to the specific attributes of transactions that markets fail to address (public good attributes, merit good attributes, externalities, and/or information asymmetries), and in a way that explicitly acknowledges the risks of government failure. Based on these two premises, managerial decentralisation/autonomy of delivery units is hypothesised to have a positive impact on efficiency.

Chapter 4 presents the methodological issues involved in the measurement of public service efficiency. These arise in three sequential stages. First, when inputs and outputs are defined and measured. Second, when the set of feasible input-output combinations is estimated (i.e. the production/efficiency function/frontier). Finally, in comparing *actual* input-output combinations with *feasible* input-output combinations. With regards to the estimation of the production frontier, research has diverged into two streams, with economists typically following the route of statistical analysis and management scientists typically opting for a non-parametric route grounded in linear programming. Four sub-sections are devoted to presenting the most important methodologies associated with each of these streams (Data Envelopment Analysis; Free Disposable Hull; and Stochastic Frontier Analysis) and discuss their relative flaws and merits.

The second half of the chapter concentrates on the specific methodological issues of measuring *public* service efficiency. Public services are normally provided free of charge (or at subsidised prices) at the point of delivery, which brings significant obstacles to the determination of their societal value and consequently to the aggregation of their output. Public service inputs normally pose fewer problems, insofar as they are purchased in competitive markets, and prices are available for determining their relative value (just as in the case of the private sector). In addition to problems in defining and measuring public service outputs and inputs, public service efficiency measurements are sensitive to the specific set of methodological choices adopted in the estimation of the productive frontier. Smith and Street (2005) lay emphasis on the consequences of four of these choices: the choice between a parametric or non-parametric approach to the estimation of the production frontier imposes distinct assumptions on output weights; decisions on the specification of the production process are likely to have an important impact on the efficiency estimates; the set of environmental variables to be included in the model, as well as the way in which these variables are controlled for is contentious irrespective of the particular approach adopted for the estimation of the frontier; finally, although allowing for dynamic effects in modelling the production process drastically increases complexity, this should be done whenever possible and the consequences of its disregard should be made explicitly from the outset.

In the second part of the thesis (chapters 5, 6, and 7) we draw on the experience of OECD countries in the providing education, health and social protection services for testing our research hypothesis. Education, health, and social protection are not only the core functions of the welfare state, but also the areas where reforms advanced the most and where the most audacious initiatives at reformulating the role of the public sector took place. The three services alone account for an average of roughly 60% of public expenditure in OECD countries (28% of GDP). Efficiency in the provision of these services is thus crucial to the overall efficiency of the public sector and to economic performance.

Chapter 5 tests our research hypothesis in the context of primary and secondary education.

Government involvement in the education sector is justified on the basis of externalities, market failures in general, and redistributive motives (Hanushek, 2002). There is a consensual sense that education involves substantial externalities (Cohn and Geske, 1990) through a variety of paths - citizen involvement in the community and government, crime reduction, family decision making and child upbringing, and economic growth. Governments may allow for these external effects and help individuals bring their decisions in line with the appropriate social calculus. Capital market and information imperfections are other types of market failures that may require government intervention. If individuals cannot barrow against their human capital, they may be forced to underinvest in education (Becker, 1993). Also, if informational problems prevent individuals from accurately assessing the influence of schools on their performance[4], they may be unable to make optimal human capital investment decisions. In both instances government interference can remove/mitigate the hindrances to efficient individual decision-making. The remaining argument for government involvement in education is usually entrenched in the desire to change the prevailing distribution of income (Juhn, Murphy and Pierce, 1993).

While these attributes may be legitimately invoked to justify government involvement in education, none of them provides for extensive public funding of education or direct public provision of education services. Externalities and capital market imperfections lead to sub-optimal investment on education. Public funding is

[4] Apart from school inputs, student performance is generally affected by a mixture of student innate ability and a range of environmental factors (namely the family socio-economic background).

therefore required to motivate individuals to investment more on education, but this is likely to be residual. Using the education system as a redistributive mechanism also implies that a certain (residual) share of the system by publicly funded. Information imperfections, on the other hand, bring the need for government action to facilitate information, but have no implications for the funding or the provision of the service.

Therefore, confining government interference to the attributes of education that call for public central coordination (externalities, market failures in general, and redistributive motives) provides for (some) government involvement in funding, but not necessarily in the provision of education services. It further suggests that when government directly provides the service through publicly owned and managed schools, these be allowed to operate on an essentially decentralised manner, in line with the incentives that guide privately managed schools.

We measure primary and secondary education efficiency through a stochastic frontier for an unbalanced data panel of 18 OECD countries in 2000 and 2003. The organisational structure of education systems is explored through two analytical components: the share of public/private providers in the system and the degree of decentralisation of decision-making procedures of public providers. The share of public providers is found to exert a negative effect on efficiency whereas the degree of decentralisation of the decision-making procedures of public providers is found to exert a positive effect on efficiency. Both results corroborate our research hypothesis.

Chapter 6 examines our research hypothesis in the context of health care services. Governments can be involved in the health sector in various ways, the most significant of which is in guaranteeing that the entire population is protected against the financial risks of sickness and medical treatment. The second most significant role is in the provision of medical services, by owning and operating medical care providers.

The way health services are financed plays a critical role in the design of their optimal mode of governance. An adequately financed health system should ensure universal financial protection. It should avoid exposing individuals to large unexpected expenses, relying as fully as possible on predictable prepayment (insurance) and minimising the share of out-of-pocket funding. It should also effectively pool the financial risks of healthy and sick individuals. Problems of moral hazard and adverse selection prevent free competitive insurance markets from

achieving these goals efficiently, which prompts extensive government intervention in the funding of health care. This result motivates our first sub-hypothesis regarding the effect of health organisational structure on efficiency: modes of governance based on direct, short-term market interactions between patients and providers, or relying mainly on unregulated private insurance are expected to be relatively inefficient.

At the service provision level, the combination of profit-maximisation with imperfect information propagates perverse incentives that lead competitive markets to failure. Public hierarchical modes of governance have the advantage of lacking a profit-maximising motif, and enabling greater political control over health service delivery, ensuring that transactions respect given priorities among interventions and patients. However, if the public sector holds the monopoly of health care provision it will lack competitive incentives for efficiency and innovation, which will tend to be aggravated by the characteristic rigidities of public hierarchical organisations reflected in the vested interests of bureaucrats and limited adaptability capacity to changing priorities and strategic orientations. This in turn leads to our second sub-hypothesis: modes of governance that combine public and private provision and that integrate market-type mechanisms into the operation of the public sector are expected to counteract the perverse incentives of pure profit maximisation, the monopolistic and bureaucratic failures of exclusive public provision, and to maximise efficiency.

Therefore, contrary to the education sector, confining government interference to the attributes of health transactions that call for public central coordination (mainly information imperfections) provides for extensive government involvement in health funding/insurance, and simultaneously significant involvement in the provision of services. On the other hand, and as in the case of education, it also suggests that public health care providers should be given sufficient autonomy to explore the benefits of decentralised coordination, and should be made to operate in an environment open to competitive forces.

OECD countries have recently introduced substantial reforms to the organisational structure of their health services, in an attempt to stimulate innovation and flexibility in the way health systems respond to political priorities. These reforms have included experimenting with market-inspired mechanisms in the internal operation of public health care providers and in their interaction with private providers. We draw on the reform experience of 23 OECD countries to test the effect of health modes of governance on efficiency. The results corroborate our two sub-

hypothesis by showing that the introduction of market-type mechanisms to public integrated health systems has a positive effect on efficiency, whereas further movement towards a market model of health care insurance and provision (public contract, and private insurance/provider models) reduces efficiency.

Chapter 7 examines the effect of modes of governance on the efficiency of social protection systems in reducing poverty.

Free social insurance markets are prone to failures: imperfect knowledge in capital and insurance markets give rise to moral hazard and adverse selection problems; general economic instability, and inflation in particular, create uncertainty about the size of future claims making it difficult/impossible to set a realistic price for the insurance; most social risks are correlated in time, which implies that the aggregate risk cannot be solved by pooling the individual risks; finally, individuals tend to be sub-optimally provident, by short-sightedness or by deliberate strategic behaviour.

Public social security systems can overcome some of these obstacles and assist individuals with the process of smoothing income across the life cycle (such as in the case of old age pensions) or across states of nature (such as in the case of unemployment benefits). Government intervention is also required at a residual level, as a means of last resort, providing a social safety net to those whose total income over the entire life-cycle is insufficient to sustain an adequate living standard[5].

The efficiency of public social protection systems is an essentially neglected issue in the academic literature. The purpose of this chapter is to estimate formally the efficiency of social protection systems and to investigate its causes. To do so we estimate a stochastic efficiency frontier for public non-pension social transfers in 19 OECD countries using panel data from the mid-1990s and the year 2000. The targeting of transfers towards the bottom deciles of the working age population is the key explanatory variable of efficiency and its effect is estimated together with efficiency itself in a single-stage procedure.

[5] In this case, income is redistributed in a way that is not actuarially linked to the contributions made to the system by the recipient.

In addition to measuring and explaining efficiency, we use our results to test a common claim of two different strands of literature[6] which suggests that there is a trade-off between the size of transfers and their targeting.

The first posits that when social transfers become very large they "tend" to be poorly targeted due to the activity of pressure groups and the progressive universalization of the entitlement to social benefits irrespective of means-testing (Tanzi and Schuknecht, 1997a, 2000). Our results do not corroborate this argument. Both the correlation between the size of social transfers and their degree of targeting towards the poor as well as the correlation between the size of social transfers and the our efficiency estimates are non significant. Hence, according to these results size is not linked to targeting or to efficiency - there is no trade-off.

The second argument asserts that the size-targeting trade-off stems from the political process. In the long run, focusing on the improvement of the targeting of transfers is said to be likely to reduce the total amount of resources available for redistribution and to have an overall negative impact on poverty alleviation (Korpi, 1983; Korpi and Palme, 1998; Mahler and Jesuit, 2004; Nelson, 2004). The size of highly targeted programs is said to be limited by the lack of political power of their main supporters (the poor), whereas social programs with extensive coverage are expected to capture support from a much more powerful political base (Korpi, 1983). Our results do not corroborate this argument either. Firstly, our evidence does not support the notion of a size-targeting trade-off. Secondly, we find no evidence of a link between the targeting of transfers and poverty alleviation.

After measuring efficiency and tracing it back to the targeting of transfers, we investigate the governance arrangements that lead different social protection systems to achieve different targeting patterns. Drawing on Kraus' (2004) organisational typology of modern social security systems, the binomial combinations Beveridge-con-social-assistance and Bismarck-without-social-assistance are found to be, respectively, the most and least efficient systems. The two policy strategies epitomise two radically different approaches to the governance of social security. The first prescribes a bureaucratic, centrally-driven remedy for dealing with a market failure. The second foresees the need for some level of central coordination in assuring a

[6] Korpi, 1983; Korpi and Palme, 1998; and Mahler and Jesuit, 2004, on the one hand, and Tanzi and Schuknecht, 1997a, 2000, on the other.

minimum level of insurance provision, but leaves the rest of the system to be coordinated by decentralised decision making procedures.

Chapter 8 draws together the thesis' main results and discusses possible future lines of research. In theory, there is an optimal *level* (and *structure*) of taxation to finance an optimal *level* of *core* public services provided in an *efficient* way. This thesis makes significant progress in comprehending the organisational drivers of public service efficiency. The next fundamental research step is to combine this knowledge with the current understanding of the optimal levels and structures of taxation and public expenditure, so that a better understanding of the optimal size and role of government may ensue.

Chapter 2: Modes of Governance

2.1 Introduction

This chapter reviews the most significant contributions of political science and economics to understanding governance and its relationship with efficiency.

The word *governance* pervades the modern anglophone social science lexicon. It is an eclectic term deployed in various contexts with diverse and often contradictory meanings. It has become a ubiquitous "buzzword" which can mean anything or nothing (Jessop, 1998). Governance can be traced back to the classical Latin and ancient Greek words for the "steering" of boats. Its original usage overlapped with *government*, the action or manner of governing, guiding, or steering conduct. For a long time usage was confined to constitutional and legal issues connected with "state affairs" and/or the direction of specific institutions or professions. In the last two decades its meaning has somewhat shifted as the word resurged in various contexts representing an autonomous concept, distinct from government. While governance would refer to the modes and manner of governing (hence the widespread use of the synonymous expression *modes of governance*), government would primarily concern the institutions and agents responsible for the governmental process, including the policy decision-making process and the subsequent implementation acts. Some authors (including Jessop) argue that the transformation of governance from a relatively dormant concept with limited scope and usage into a fashionable analytical device is linked to a global paradigmatic crisis in the social sciences in the 1970s and 1980s – an increasing dissatisfaction with the capacity of prevailing theoretical frameworks to describe and explain the empirical world. At a more pragmatic level, other authors indicate the crisis of the welfare state and the failure of reform policies during the 1970s as a key background factor for the emergence of the academic notion of governance and its early emphasis on new models of political steering (Mayntz, 2003).

2.2 Modes of Governance and Political Science

Governance started its scientific terminological career in the fields of economics and organisational theory, from where it found its way into public

management and political science (Mayntz, 2003; 2004). In recent years there has been a growing trend in public management and political science literatures towards analysis in terms of multiple modes of governance (Tenbensel, 2005). This trend is in part a reaction to the increasingly sterile ideological debate between old public administration and new public management. The debate is in itself a circumscribed version of a broader argument on the relative merits of markets and hierarchies as modes of social and economic coordination, imported in the 1980s and 1990s from its usual macro setting and brought to the discussion of modes of organising and governing the public sector.

Political scientists have been actively involved in the conceptualisation of governance (for a recent overview, see Kersbergen and Waarden, 2004; Kooiman, 2003; Mayntz, 2004; Pierre, 2000a). Although the terms *governance* and *modes of governance* have been broadly used to refer to the *steering and coordination of interdependent actors based on institutionalised rule systems* (Benz, 2004), the diversity of specific analytical meanings that pervades the literature is paramount. A wide variety of different phenomena are associated with these terms, ranging from a multitude of institutional structures and actor constellations in political decision-making, to various types of policy instruments. Rosenau (2000, p.171) describes governance as "systems of rules, as the purposive activities of any collectivity, that sustain mechanisms designed to ensure its safety, prosperity, coherence, stability and continuance". For Teib *et al.* (2004, p.1-2), "a workable definition of governance is best attained if we concentrate on the policy dimension", and so governance is defined as "political steering, and modes of governance denote different styles or instruments of political steering". Jachtenfuchs and Kohler-Koch (2004, p.99-100) define governance as "the continuous political process of setting explicit goals for society and intervening in it in order to achieve these goals". These three definitions highlight three different perspectives on governance: as a system, a process, or a set of activities (Diedrichs, 2004). However, these are merely representative examples of the ample variety of alternative perspectives that characterises the political science literature on governance.

In addition to this plurality of meanings and connotations, the scope of the term is also subject to different interpretations. In a broad understanding, governance is related to different forms and ways of taking decisions for and within society as a whole, i.e. a "pervasive form of political steering" (Mayntz, 2003). From a narrow

point of view, governance is conceived as a counter-model to the ideal-type "nation state government" based on hierarchical steering of society through authoritative instruments and hinging on a relatively homogeneous source of legitimacy represented by the national electorate (Mayntz, 2004). The rising role played by international organisations, sub-national institutions and other non-state actors in defining the overall path of society has generated increasing scepticism on the idea of the nation state as the main political actor and its ability to steer society along traditional hierarchical lines (Rosenau and Czempiel, 1992). Growing attention has been devoted to the conceptualisation of different types of governance arrangements where coordination rather than hierarchical steering prevails. These arrangements are the exclusive attributes of governance when it is narrowly defined (Diedrichs, 2004).

In a recent note towards the conceptual clarification of *modes of governance* Treib *et al.* (2005) argue that despite a common concern for the relationship between state intervention and societal autonomy, different strands of literature focus on different facets of this continuum: the politics, polity and policy dimensions of governance.

In the politics dimension, governance is related to the process of policy-making. It is about the ways in which the divergent preferences of citizens are translated into effective policy choices, the ways the plurality of societal interests are transformed into unitary action and the compliance of social actors is achieved (Kohler-Koch, 1999). In this context, the critical distinctive feature of different modes of governance is the relationship between public and private actors in the policy-making process. At one extreme, a hierarchical state leaves the policy process to public actors. At the opposite extreme, self-regulations by firms without state intervention or self-organisation of communities only involves private actors. Between these extremes there are various intermediate modes of governance involving public and private actors. Typical examples include several forms of *policy networks* (Marin and Mayntz, 1991), and bureaucracies restructured along New Public Management lines, transferring formerly public tasks to the private sector (Kersbergen and Waarden, 2004; Mol, Liefferink, and Lauber, 2000). At the empirical level, only hybrid modes of governance including both public and private actors are said to exist. The state is dominated by public actors, whereas markets are dominated by private actors (Streeck and Schmitter, 1985). But even processes of private self-

regulation may be influenced by the state, as their own existence hinges on the threat of state coercion (Héritier, 2002; Mayntz and Scharpf, 1995).

In the polity dimension, governance is seen as an institutionalised system of rules that shapes the actions of social actors (Rosenau, 1992; Mayntz, 2004). Different modes of governance are placed on a continuum defined by "markets" and "hierarchies" as the two ideal poles. Irrespective of the type of actors (public or private) involved in the decision-making process, the institutional structure governing their interactions can either be hierarchical, allowing some actors to achieve collectively binding decisions without the consent of others, or it can be inspired on a market structure, allowing each actor to voluntarily decide on their course of action. In between these opposing extremes, various other institutional structures coexist (e.g. *communities*, *associations*, and *networks*, Schneider and Kenis, 1996), the network structure being perhaps the most notorious - a non-hierarchical constellation of interdependent actors linked by different types of power resources. Again, this categorisation of institutional arrangements is seen as a theoretical abstraction. Empirically only hybrid forms involving elements of several ideal-type modes of governance can be found. Effective steering and coordination can only be accomplished though the practical combination of these elements (e.g. markets rely on hierarchical authority to enforce contracts, Streeck and Schmitter, 1985). An institutional feature that often stands out in the polity dimension of governance is the degree of centralisation/dispersion of the locus of authority, both in a horizontal (among central state actors) and vertical (among territorial units or boundaries of the state) perspective (Walker, 2005; Grosse, 2005).

In the policy dimension, governance is defined as a *mode of political steering* (Héritier, 2002). Different modes of governance refer to the employment of different steering instruments for the achievement of particular policy goals. The state can intervene through various types of more or less heavy-handed instruments to achieve given societal outcomes: command and control; incentive and supply, information, deliberation and persuasion, as well as forms of social influence and control (Baldwin and Cave, 1999; Windhoff-Héritier, 1987).

Although the politics, polity and policy dimensions of the governance debate carry with them specific analytical focuses and highlight particular facets of the governance phenomena, they share in common the discussion of the role of the state in society (Pierre, 2000b) by looking at aspects of the relationship between state

intervention and societal autonomy as a continuum from public authority to societal self-regulation (Treib *et al.*, 2005). Additionally, they share in common the separation between "ideal" conceptual modes of governance and "real" empirical modes of governance, emphasising the hybridity of these later.

2.3 Modes of Governance and Economics

The conceptualisation of *modes of governance* as forms through which order is achieved and sustained in economic and social life has been the subject of intense debate for many years and lies in the centre of economic analysis. The establishment of a *market-hierarchy* dichotomy as the two building blocks of economic organisation is as old as economic science itself. The discussion of the relative merits of a market driven economic system as opposed to a central planned economy can be traced back to Smith (1976) in his foundational work, and is in itself a discussion about the properties and capabilities of markets and hierarchies to produce co-ordination at the macro level.

However, at the micro-analytical level it took a long time for economists to recognise the existence of alternative methods of economic coordination. Notwithstanding the early efforts of some prominent members of the profession in drawing attention to the peculiarities of business organisations in coordinating production (e.g. various writings from Marshall; Clark, Knight, and Robertson[7]), standard microeconomics relied for decades on the concept of firms as production functions, i.e. a set of technologically determined input-output combinations. During that time, the study of organisation developed in other disciplines to the point of generating a field of its own (Organisation Theory).

The development of a theoretical framework for identifying and characterising alternative modes of organising transactions and investigating potential trade-offs between them was accomplished at the margin of neoclassic orthodoxy by New Institutional Economics[8].

[7] The author is responsible for a famous metaphor of business organisations in referring to them as "islands of conscious power in [an] ocean of unconscious co-operation" (quoted by Coase, 1937, p. 333).

[8] The term "New Institutional Economics" (NIE) was first coined by Oliver Williamson in 1975 and refers to a strand of economic literature that has gained increasing influence over scholarly research in the last three decades, not only in economics, but also in legal studies, political science, sociology, anthropology, management sciences, and other fields. In a recent review of the topic, Ménard and

The seeds of the NIE governance approach were set out by a series of promising ideas advanced, mainly independently, in the 1930s. Following the contributions of Karl Llewellyn (1931) and John R. Commons (1932), Ronald Coase's groundbreaking article on "The Nature of the Firm" (1937) established a turning point in the economic inquiry of organisation[9]. Almost simultaneously, Chester Barnard (1938) reflected on the "marvel of hierarchy" publishing "The Functions of the Executive" and emphasising the role of "authority" in demarcating firms from markets. Some of the pioneering works of subsequent decades included Friederich Hayek's (1945) reflections on "the marvel of the market" and in particular the coordinating efficiency of the price mechanism; Simon (1951) on the employment relationship; and Arrow (1964) on the role of control in hierarchies.

The prevailing NIE model of economic organisation originates from the 1970s through several publications that built on these preliminaries. In a draft version of his future theory, Williamson (1971) focused on transaction costs in examining "vertical integration" and stressed the importance of contracts as a key organisational device. Alchian and Demsetz (1972) followed with a controversial re-examination of the Coasian approach and interpreting the firm as a nexus of contracts. Arrow (1974) also made is contribution to further extending the importance of organisational issues on the economics agenda. In 1975, with the publication of *Markets and Hierarchies* Williamson established a new turning point in the NIE movement, assembling "dispersed elements (including his own contributions) into a coherent framework that linked transaction costs, contractual arrangements, and modes of organisation, thus

---

Shirley (2005, p. 1-2) characterise NIE as the study of "institutions and how institutions interact with organisational arrangements. Institutions are the written and unwritten rules, norms and constraints that humans devise to reduce uncertainty and control their environment. These include (i) written rules and agreements that govern contractual relations and corporate governance, (ii) constitutions, laws and rules that govern politics, government, finance, and society more broadly, and (iii) unwritten codes of conduct, norms of behaviour, and beliefs. Organisational arrangements are the different modes of governance that agents implement to support production and exchange. These include (i) markets, firms, and the various combinations of forms that economic actors develop to facilitate transactions and (ii) contractual agreements that provide a framework for organizing activities, as well as (iii) the behavioural traits that underlie the arrangements chosen".

[9] Coase's article, although left aside for more than thirty years after it was originally published, went on to become a major driving force to the evolution of economic thought on organisational issues. Terry Moe (1984, p. 743) identifies three notions that have proved particularly influential: "(1) Economic organisations are best understood by comparing their efficiency to that of the market. (2) In the real world, which is clearly not characterized by perfect competition, perfect information, or frictionless exchange, economic activities and organisational arrangements are best understood in terms of the transaction costs inherent in any system of exchange relationships among rational individuals. (3) These relationships are contractual in nature, and the firm is best understood as founded upon a distinct kind of contractual arrangement, the authority relation".

providing a model that remains at the core of the micro-analytical branch of NIE" (Ménard, 2005, p. 283). At the end of the decade, Klein *et al.* (1978) examined the role of specific investments and risks of hold-up as the explanation to the choice of a mode of organisation.

Transaction cost economics (TCE) and property rights theory, both rooted in the new institutional approach, are the most traditional and well established economic theories of organisation. However, they are not the only existing economic explanations of organisation as they now compete with a set of resource-based/evolutionary perspectives of the firm. Serious divergences divide all these alternative explanations, and an integrated economic theory of organisation is still missing. In the following paragraphs we address the key features of some of the main approaches.

2.3.1 Transaction Cost Economics and Property Rights Theory

The most conventional approach to economic organisation sees governance as "the means by which to infuse order, thereby to mitigate contractual hazards and realize mutual gains from trade…[i.e.] the deployment of alternative modes of governance among differing transactions with the purpose of economizing on transaction costs" (Williamson, 2004, p.0). Modes of governance in turn are understood as institutional arrangements within which a transaction or set of related transactions are decided upon and implemented (Ménard, 2005)[10].

TCE focus on the microanalytics of transactions on the one hand, and of alternative modes of governance on the other. It analyzes and compares the competence of different modes of processing and monitoring transactions[11] to take advantage of the division of labour while simultaneously economising on bounded rationality and safeguarding parties against contractual hazards.

All modes of governing transactions involve costs. The sources of these costs are to be found in (i) the propensity of economic agents to behave opportunistically which generates the risk of contractual hazards and the associated need to define and

[10] Recently Williamson defined a governance structure as "the institutional matrix in which the integrity of the transaction is decided" (Williamson, 1996, p.378).
[11] Coase uses the expression "institutional structure of production" to define these arrangements, whereas Williamson speaks of "mechanisms of governance".

implement costly safeguards; and (ii) in certain types of noncontractible uncertainties that plague the environment where transactions take place.

In comparing the relative advantages of alternative ways of organising transactions, TCE first focuses on the attributes of transactions thought to determine variations in their costs. The most important of these attributes are the specificity of assets involved, the uncertainties surrounding the transaction at stake; and the frequency with which the transaction occurs (Williamson, 1985).

Asset specificity[12] creates mutual dependence and opens the possibility of "hold-up" problems by exposing the owners of specialized assets to the opportunism of trading partners owning other specialized and complementary assets who may seek to capture inappropriate quasi-rents stemming from the combined use of such assets (Araujo *et al.*, 2003).

Significant levels of uncertainty and complexity entailed by the transaction and high degree of difficulty in measuring performance pose important obstacles to the construction of mutually satisfactory agreements, forcing the parties to rely on contracts that merely specify rights, obligations and procedures rather than actual performance standards. In extreme cases this solution may itself be insufficient and internalising those transactions previously obtained through the market may become a better or indeed the only viable alternative.

Finally, the frequency with which similar transactions occur and the duration or period of time over which they recurrently take place is another basic attribute. As a general principle, high frequency transactions taking place over long periods of time tend to be internalised within formal organisations or to be managed under specially created low-cost routines that reduce the need for formal mechanisms to enforce agreements between the parties.

Williamson´s (1985) discriminating alignment hypothesis establishes the link between transaction costs and modes of organisation: calculative agents operating in a competitive environment will adopt the mode of organisation that fits comparatively better with the attributes of the transaction at stake. This hypothesis effectively enables TCE to circumvent the study of the inner-workings of alternative modes of governance responsible for their comparative ability to organise transactions efficiently.

[12] The value of investments that would be lost in any alternative use.

Property rights theory (PRT) pioneered by Grossman and Hart (1986) and Hart and Moore (1990) shares with TCE the importance ascribed to the role of incentives in aligning the interests of cooperating parties in a context of transaction costs and incomplete contracts. In both approaches, the design of incentive schemes and ownership patterns that support those incentives are the key levers of the efficient organisation of transactions.

Similarities apart, the two approaches differ at least in three important ways (Whinston, 2003). First, from a methodological point of view, PRT is substantially more formal than TCE. Second, PRT explicitly concentrates on distortions in ex-ante investments, whereas TCE's main analytical focus is on ex-post haggling and maladaptation costs. Third, while TCE assumes that opportunism can be mitigated if the transaction is coordinated through hierarchical command (at the expense of bureaucratic costs), PRT assumes that this risk is present in all forms of organisation as investment and trading decisions are thought to be decentralised regardless of the structure of asset ownership.

An important advantage of PRT is that it investigates the costs and benefits of integration in a manner that does not rely on the presence of an impersonal market as TCE does. A firm is seen as a set of non-human assets under common ownership. Control over assets gives the owner bargaining power when unforeseen or uncovered contingencies force parties to negotiate how their relationship should be continued. The owner of an asset can decide how it should be used and by whom, subject only to the constraints of the law and the obligations implied by specific contracts. Assets become bargaining levers that influence the terms of new agreements and hence the future payoffs from investing in the relationship. In contrast to TCE, PRT models assume that all bargaining, including any that occurs after investments are made, is efficient. Thus, everything turns on how ownership affects initial investments (Holmström and Roberts, 1998).

2.3.2 Resource-based/Evolutionary Theories

The conventional conception of economic organisation (including TCE and formal property rights theory) regards the emergence of firms as an instance of market failure. Firms are seen as legal, contractual entities, owners of physical assets providing investment incentives and facilitating the resolution of hold-up problems.

The capabilities perspective, resource or competence based view of the firm is an alternative conception of economic organisation that sees the problem of economic coordination as an attempt to find integration mechanisms that sustain the division of labour amongst agents with incomplete, dispersed and disparate knowledge and help create and test new knowledge (Piore, 1992; Loasby, 1998a, 1998b; Kogut, 2000). Firms are conceived of as a bundle of essentially tacit "capabilities", shared by the human assets that comprise it. The boundaries of the firm are "partly determined by the differential between the capabilities they hold in relation to potential supplies and the relative costs of developing capabilities internally against accessing them via non-hierarchical means" (Araujo *et al.*, 2003, p.1263). Opportunism and threats of hold-up are part of the calculus of deciding which is the most cost-effective way to access capabilities (Barney, 1999).

The writings of Penrose (1959), Richardson (1972) and Loasby (1998a, 1998b) are amongst the main contributions to the development of the capabilities approach to economic organisation.

Penrose's (1959) underpinned the division of labour and the growth of the firm on a distinction between resources and the services they render. Resources are capable of providing a variety of productive services which, in turn, are capable of modifying the attributes of resources and thus enable the provision of new services. The disposal of productive resources between different uses and over time is determined by administrative decision. The boundaries of the firm are thus defined by the nature of its managerial and administrative responsibilities (Penrose, 1995, p.xvi).

Richardson (1972) uses the notion of capabilities and activities instead while addressing issues of division of labour and coordination of capabilities in industrial settings. For Richardson the boundaries of the firm are delineated by the type of activities it performs, the integration of these activities with others and the capabilities underpinning these activities: "...organisations will tend to specialise in activities for which their capabilities offer some comparative advantage; these activities will, in other words, generally be similar...although they may nevertheless lead the firm into a variety of markets and a variety of product lines" (Richardson, 1972, p.891). The grouping of activities under hierarchical coordination is explained by interrelationships between activities and associations between the capabilities that support them.

Complementary activities have to be matched in level and/or specification and require sequential coordination. Close complementary activities are to a certain extent specialized to each other and require more complex forms of coordination in the way they are combined (Richardson, 1972, p. 891). As for the resources that support them, activities can be classified as similar or dissimilar depending on the similarity of the capabilities they use.

Different combinations of similarity and complementarity favour different modes of economic coordination. Complementary but dissimilar activities may be efficiently coordinated through market exchange without any form of ex-ante planning. Close complementary activities require concurrent coordination via intra or interfirm cooperation, where the sequential adjustment of activities is consciously planned or put together through the articulation of a limited number of decision making units. Close complementary and similarity activities favour intrafirm cooperation. Close complementarity and dissimilar activities favour interfirm cooperation as the coordination of dissimilar capabilities may prove impossible under a single administrative framework. Firms are here seen as institutional devices enabling the organisation of a set of closely complementary and similar activities that have to be put in place simultaneously and thus require conscious, planned direction (Richardson, 1998). The distinct nature of firms is "not simply the professional knowledge and skill its members bring with them but the complex pattern of relationships and 'local knowledge' derived from teamwork and continued interaction" (Araujo, *et al.*, 2003, p.1262).

Loasby (1998a, 1998b) developed Penrose and Richardson's arguments by relating the definition of capabilities to the "knowledge-how" required for productive activities. "Know-how" is the domain of skilled performance, learned through situated practice and emulation of experienced performers (Araujo *et al. 2003*), and it can either be direct or indirect depending on whether the firm directly controls the capabilities itself or may have access to them in an organised way.

Although resource-based/evolutionary theories of the firm are becoming ever more influential, the most well established approaches to economic organisation are still rooted in NIE (most notably transaction costs economics). NIE deviates from the standard neoclassical assumptions regarding perfect information, unbounded rationality, and costless/instantaneous transactions. Incomplete information and

limited mental capacity makes individuals face uncertainty about future events and incur transaction costs. To reduce risk and transaction costs individuals create institutions and develop modes of organisation embedded in those institutions (Ménard and Shirley, 2005). Markets, firms and hybrids are the archetypes of these modes of organisation. Our analysis now turns to the characterisation that NIE-inspired literature has made of these institutional arrangements.

2.3.3 Markets

Market can either designate a mode of organising economic transactions, alternative to firms and hybrid organisations, or the general set of arrangements that characterise a market economy, i.e. the set of institutions that embed all modes of organisation[13].

As a mode of organisation/governance market is a sub-class of the various existing institutional arrangements for transferring rights. Its defining trait consists in the exchange of property rights through mechanisms that require mutual consent of the parties involved and that coordinate decentralised decisions made by these parties using information provided by the price system (Coase, 2005).

While NIE and neoclassic economics have in common the importance accorded to the role played by the price mechanism in coordinating transactions, their views diverge with regards to several other significant issues: NIE does not conceptualise markets as pure structures, but rather as entities shaped by institutional factors; for new institutional economists prices do not coordinate, but rather send signals to those coordinating; furthermore, agents do not adapt passively to prices, but instead carry out all sort of strategic actions that actively affect them.

The acceptation of these premises has two important implications. First, markets require specific institutional supports to emerge and develop. These supports combine legal, political, and social factors, with enforcement of contracts as a fundamental element (White, 1981; North, 1981; 1990) alongside the definition and implementation of property rights (Alston, Libecap and Mueller, 1997).

[13] Ménard (2005, p. 302) refers to this particular interpretation of market as "the central economic institution in last resort in that at some point all modes of organisation intersect with and/or are embedded in markets, e.g. firms and hybrids obtain resources through voluntary exchanges, compete in capital and labour markets, etc."

Second, since markets are institutionally embedded and shaped by formal and informal rules, they differ according to the arrangements that support them. Although they may share some common properties, namely those captured by microeconomics through the analytics of supply and demand, the heterogeneity of their institutional underpinnings influences the comparative costs and benefits of using them. The practical reflection of this influence is the potential variety of mechanisms involved in the formation of prices[14], entailing different processes, arrangements and rules and translating into different transaction costs.

2.3.4 Firms

In the neoclassical realm the firm is "a unitary profit-maximising entity defined by a technologically determined production function (Yarbrough and Yarbrough, 1988, p.2). The embedding models assume that monitoring is costless, or can be inexpensively endogenised through an adequate contract; shirking can be detected and punished on the basis of perfect revelation of information and no enforcement problems; and employees are subject to a fully contingent contract.

NIE addresses integrated "formal organisations" (Barnard, 1938) primarily as governance structures. It recognises the importance of technology in defining the set of feasible activities, but it subsumes this production function dimension of firms under the notion of governance structure. This in turn is depicted as a combination of legal, economic and social dimensions. As legal entities, firms operate and are liable as individual agents capable of transferring rights. As economic devices, they rely on a multitude of contractual arrangements structured along hierarchical lines to coordinate transactions. As social units, firms have motivational properties shaping their members' behaviour that go beyond monetary incentives. Although these different dimensions are equally important, NIE focuses mainly on the economic component, studying the specific properties that differentiate firms from other organisational arrangements and that confer them an edge in organising some transactions.

The main assertion is that the comparative advantage of firms derives from their capacity to organise some transactions through centralised command in a cost effective way when compared to decentralised coordination via the price system

[14] Posted prices, prices determined by auction, prices formed through negotiation, etc.

(Coase, 1937). The distinctive feature of these hierarchical "system[s] of consciously coordinated personal activities or forces" (Barnard, 1938, p.72) is thus command – a relationship in which those who are assigned the implementation of a particular task have to report to the persons in charge who, in turn, can be accountable for the performance of that task (Ménard, 2005).

A variety of arguments have been advanced to justify the comparative advantage that can be expected from these hierarchical relationships. It is thought that the ability of supervisors to reallocate human resources without negotiation reduces transaction costs and mitigates or eliminates some sources of uncertainty (Simon, 1951; Beckmann, 1988). The internalisation of transactions is said to provide the means for the extension of the domain of rationality, improving the quality of decisions through division of "cognitive labour" (Aoki, 2001). Furthermore, it is frequently argued that the communication system supervised by the entrepreneur can be comparatively efficient when the information provided by the market is costly, incomplete and difficult to process (Alchian and Woodward, 1987).

These potential advantages of hierarchical modes of governance need to be weighted together with the administrative/bureaucratic costs they create before a complete judgement of their relative merits in coordinating transactions can be made.

Elements of both the comparative advantages and costs of hierarchies over alternative modes of coordination can be found in the three essential components of command: control, cooperation, and communication.

Control makes command credible by providing the means for implementing orders, evaluating the adequacy of actions chosen, and checking on the propensity of members to renege their commitments (Williamson, 1985; Beckmann, 1988; Demsetz, 1995). Control provides flexibility, allowing supervisors to monitor the conformity of action to orders and, if need be, reallocate tasks without renegotiating contracts and using the price system. Under certain circumstances, this endows hierarchies with the capacity to re-adapt the organisation of transactions faster then decentralised coordination (Bolton and Farrell, 1990). Control provides powerful tools to constrain opportunism through interactions among levels of management (Tirole, 1986). Some authors have recently suggested that central control may facilitate the conduction of "controlled experiments" and help determine how assets can be organised more effectively (Foss *et al.*, 2002). Finally, control allows settling disputes faster and less costly than arbitration by third parties (Williamson, 1975;

Dow, 1987), and is often superior to external control in collecting/processing the relevant information and making the necessary adaptations (Williamson, 1975).

New institutional economists have been keen on showing the positive side of control, but they have also gone the other end of the field calling attention to the rigidities and costs that draw a limit on the efficiency of centralised coordination. Demsetz (1988b, 1995, 2002) pinpointed the costs of excluding non-owners from using resources as a major limitation to control in large organisations. Hansmann (1988) demonstrated that owners of physical assets are as much concerned with controlling the use of their assets as with controlling residual profits, which imposes significant costs to hierarchies when compared to market coordination. Williamson (1985) identified the non-replicativity within firms of market incentives as the main obstacle to lessening administrative costs. Milgrom and Roberts (1990; 1992) disclosed the costs arising from influence activities among managers and from the loss of information along the numerous tiers that characterise control in hierarchical structures.

Apart from control, cooperation is another vital component for the exercise of command (Simon, 1991). Although subject of controversy, cooperation can generally be defined as "the willingness of agents to pool resources even when they cannot assess ex-ante the benefits expected, or if there are any benefits at all to be expected in doing so" (Ménard, 2005, p.292). The presence and significance of cooperation for the support of hierarchies has long been recognised in the literature. Alchian *et al.* (1987, p.1031) characterised the firm as "a contractually related collection of resources of various cooperative owners" and "the organisation of cooperative joint production". Following Common´s (1934) valuation of mutuality and Arrow's (1974) emphasis on the economic value of social interactions, Williamson (1975) explored how cooperation, in the form of "attitudinal interactions", can mitigate the costs of control in formal organisations by alleviating the risk of conflicts and enhancing their capacity to settle disputes[15].

As with control, cooperation is also open to limits and costs. Williamson (1975) identified some of the sources: free riding strategies through ex-ante selection of members and malingering behaviour once selected (ex-post); collective decision-

[15] Williamson (1975) identifies four main advantages deriving from a cooperative "atmosphere": (i) scale economies in the acquisition of information; (ii) dispersion of risk through the different members of the group in the event of unanticipated contingencies; (iii) mitigation of moral hazard and adverse selection; (iv) increased productivity due to a better developed "sense of responsibility".

making detrimental to the advantages of command; incentives to collude and develop side-payments; and high costs of processing information and communicating in a team oriented organisation.

Together with control and cooperation, information and communication completes the triptych of command. A key organisational feature of hierarchies when compared to other modes of governance is their ability to gain advantage in processing and communicating information under particular circumstances. Formal organisations can build and develop routines that make codification possible and reduce internal costs. The construction of a common language within the firm (corporate culture) facilitates the sharing of knowledge. The combination of human resources extends the capacity of individuals to absorb knowledge while the organisation of transactions by command promotes an efficient processing of information, its fast transformation into action and consequently rapid adaptation (Williamson, 1975; Aoki, 1986; Demsetz, 1988a, 1988b, 1995).

As with the other two elements of the triptych, the informational structure underpinning internal coordination is not exempt from costs either. Hierarchies have a tendency to generate informational noises and uncertainties of their own which pose serious limitations to the comparative efficiency of command. Some contributions in the literature have addressed these limitations. Williamson (1967) examined how a small noise in the transmission of signals in multi-layered hierarchies imposes limits on the size of formal organisations. Demsetz (1988b, 1995) explored the "decreasing returns" to the capabilities of business managers to monitor information. Aoki (1986; 1990) showed the trade-off in processing information between centralised organisation (accumulating noises along the various hierarchical layers) and decentralised coordination of dispersed information.

2.3.5 Hybrid Arrangements

NIE initially concentrated on the study of hierarchy as an alternative to markets in coordinating transactions. Other modes of governance falling in between these two extremes were deemed unstable and transitory. Until the mid-eighties a limited number of articles were published on inter-firm contracts (Klein *et al.*, 1978; Ouchi, 1980; Eccles, 1981; Cheung, 1983), franchising (Rubin, 1978), and "non-standard contracting" (Williamson, 1975). A turning point arose with the

transformation of transaction costs economic into an empirical research program (e.g. Williamson, 1985; Palay, 1984; Masten, 1984; Joskow, 1985). In the years that followed, "hybrid" forms of organisation (Williamson, 1991) became the core topic of a burgeoning literature initially based on non-economic journals.

The terminology contained in these studies is unstable. "Hybrids", "clusters", "networks", "symbiotic arrangements", and "chain systems" are often used interchangeably. The underlying organisational forms are equally heterogeneous: subcontracting, networks, alliances, franchising, collective trademarks, partnerships and cooperatives[16]. Nonetheless, there is a common thread uniting these concepts. All refer to forms of inter-organisational collaboration whereby distinct property right owners operating distinct legal entities, coordinate sub-set of transactions (share or exchange technologies, capital, products, and services) through specific, mutually agreed forms of governance (Ménard, 2005).

The literature on these forms of inter-organisational collaboration reveals several key regularities that define their distinctive character. First, hybrids are created with the purpose of sharing some resources, coordinating some decisions and generating rents. Joint organisation of activities based on inter-firm coordination enables the pooling of resources and capabilities that would not have been adequately bundled through markets (Teece and Pisano, 1994), while hierarchical coordination would reduce flexibility, create irreversibility and deteriorate incentives. However, pooling resources can also be a source of conflicts and costs. The process of rent distribution is often controversial and can destabilize the organisational arrangement. Also, pooling resources requires continuity in the relationship and cooperation, which in turn impose limitations on the autonomy members benefited under a market arrangement, without the compensation of control that hierarchies concede.

A second regularity of hybrids is the peculiarity of the structure and purpose of their underpinning contracts. Hybrids rest on relational contracts, i.e. frameworks for "transactional reciprocity" (Park, 1996) linking activities and resources among members who simultaneously operate unrelated transactions. The relational dimension is ingrained in the advantages and risks of sharing resources between

[16] Some of the most significant references include (Ménard, 2005, p.295): (i) on subcontracting: Eccles, 1981; Aoki, 1988; and Bajari and Tadelis, 2001; (ii) on networks; Thorelli, 1986; Powell, 1990; Podolny and Page, 1998; (iii) on alliances: Oxley, 1999; Baker *et al.*, 2003; (iv) on franchising: Rubin, 1978; Williamson, 1985; Lafontaine and Slade, 1997; (v) on collective trademarks: Dwyer and Oh, 1988; Ménard, 1996; Sauvée, 2002; (vi) on partnership: Farrell and Scotchmer, 1988; Powell, 1996; and (vii) on cooperatives: Cook, 1995; Cook and Iliopoulos, 2000.

legally independent actors (Goldberg, 1980; Williamson, 1985; Baker *et al.*, 2002). Advantages include increased market shares, transfer of competencies and access to financial resources. Risks emerge from partners coordinating only part of their decisions, subject to unforeseeable revisions depending on the level of uncertainty associated with the processes, products or demand targeted by the joint investment.

The relational dimension of hybrids helps fill the gap left by incomplete contracts, monitor partners, and solve conflicts minimising repeated renegotiations. It also helps to economise on the costs of extensive contracting among autonomous partners, and thus maintain some advantages over the administration costs of single integrated hierarchies. This could not be achieved though without the intervention of other complementary internal governance mechanisms. That intervention is usually credited on formal governing bodies, which are made responsible for subclasses of decisions transferred from partners with a view to economising on transaction costs and stabilising relations (Ménard, 1994, 1996, 1997, 2004). Empirical studies suggest high volatility in the degree of formalism and power conferred to these governing entities, which is thought to reflect differences in the significance of contractual hazards and associated transaction costs. Ménard (2005, p.301) proposes a four-category typology:

***

At one end of the spectrum, close to market arrangements, hybrids rely primarily on *trust*: decisions are decentralised and coordination relies on mutual "influence" and reciprocity. At the other end, hybrids come close to integration, with tight coordination through quasi-autonomous *government bodies* or "bureaus" sharing some attributes of a hierarchy. Between these polar extremes, mild forms of "authority" develop, based on relational networks and on leadership. Relational *networks* have attracted a lot of attention in organisation studies (Powell, 1990; Hakansson and Johanson, 1993; Grandori and Soda, 1995). They rely on tighter coordination than trust, with formal rules and conventions based on log-term partnerships, on complementary competences, and or on social "connivance" (Powell, 1996). By contrast, hybrids coordinated by a *leader* leave little room for autonomy although some formal symmetry can be maintained. Subcontracting, particularly with long-term contractual relationships, or alliances related to R&D projects are often of that mode (Eccles, 1981; Pisano, 1990; Powell, 1996).

***

A third regularity of hybrids is the presence of competition. Hybrids involve complex combinations of interdependence and autonomy in which competition arises

at two different levels. One the one hand, within these organisational arrangements partners cooperate on some issues, but they also compete against each other. Bilateral agreements with long-term contracts are often subject to internal competition since individual strategies remain autonomous (Coase, 2000). Collaborative agreements can be purposively designed so as to encourage recurrent competition as in subcontracting (Eccles, 1981; Dyer, 1997). Activities may overlap as partners compete to attract costumers from the same sub-set (Raynaud, 1997). Partners may cooperate on some activities and compete on others (Baker *et al.*, 2003). On the other hand, outside the sphere of the collaborative organisational arrangement, competition also exists amongst hybrids and other arrangements through the market.

Features of these regularities can also be found in markets and hierarchies. Therefore, the distinctive character of hybrids is not so much to do with the presence of any of these individual regularities, but rather with the specific mix of competition and cooperation that subordinates the role of prices in markets, and command in hierarchies (Jorde and Teece, 1989; Grandori and Soda, 1995; Ménard, 1997). This mix is the creator of a distinctive mode of governance.

The particular incentives that have boosted the creation of these arrangements have been explored by Williamson (1991) based on the model originally applied to the market-hierarchy dichotomy. The fundamental idea is that when investments are specific enough to generate substantial contractual hazards, without justifying integration and its costs, and when uncertainties are consequential enough to require tighter coordination than that provided by markets, there will be strong incentives for the constitutions of hybrid modes of organisation. Uncertainty is secondary to specific investments in that without asset specificity the parties would transact through the market and there would be no hybrid. However, investment-specific relationships tend to be infused with uncertainties about the level of resources pooled and their monitoring. In practice, it is the combination of investment specificity and uncertainty that provides the proper incentives for the creation of hybrids, whereas the presence of only one of these attributes leans towards contract-based arrangements. In the presence of consequential uncertainty, hybrids need to combine flexible adaptation capacity, control, and safeguards against opportunistic behaviour. The intensity required for these items determines the degree of centralisation in the governance of hybrids. The more consequential uncertainties are the more centralised coordination will tend to be (Ménard, 1996; 1997; Nootebom, 1999).

2.4 Friederich Von Hayek on Markets and Hierarchies

Transaction costs economics, arguably the most influential economic approach to organisation, treats market trade as a default that is assumed superior to within-organisation trade unless levels of uncertainty, frequency and asset specificity are high enough to pull the transaction out of the market. Because the market is the default, its benefits are not spelled out as clearly as its costs. The market is treated as a black box, much in the same way as neoclassical microeconomic theory treats the firm. On the other hand, a variety of conditions have been adduced by Williamson and others to limit firm size - costs of bureaucracy, the weakening of individual incentives, the hazards of internal politicking, and so on - but none of these costs are fully understood either.

In order to overcome the analytical limitations of TCE and NIE theories of organisation at large, we argue in the next chapter that economic organisation is essentially about bringing order into systems of human interaction through a multitude of modes of governance, ranging from conscious deliberate design to spontaneous self-reinforcing rules. The effects of each mode of governance on efficiency depend on the nature and scale of the transactions upon which they are deployed. Both assertions follow closely Friederich Hayek's reflections on the inner workings of markets and hierarchies as modes of coordination and his seminal distinction between *spontaneous self generating order* and *organized ("made") order*. The following paragraphs highlight the importance of these concepts as the fundamental pillars of economic coordination with potential to reunite and expand NIE and resource-based/evolutionary theories of organisation.

The importance of Hayek's theoretical framework to the economics of organisation, and in particular its potential for establishing a bridge between various theoretical strands has recently been captured by Ioannades (2003, p.533, 534, 535):

***

According to the standard periodization of Hayek's work, we have to distinguish between early Hayek, the analytical economist, and the mature Hayek, who focused on wider spheres of the social sciences like social philosophy, political theory, and cognitive psychology. Taken literally, this periodization seems to imply that it is his early period that may be of relevance for today's analytical

economics. Contrary to this view, we will argue that contemporary economics has a lot to learn even from that part of Hayek's work that is not purely economic.

We focus here on Hayek's distinction between two types of social order: spontaneous orders and organisations…this distinction offers an important starting point for a theory of economic organisation…[it] allows us to construct an evolutionary approach to the firm…it introduces an understanding of the business firm as a process, in contrast to the static outlook of contractarian theories, while still allowing us to describe the firm as the outcome of a contracting process among asset owners. The importance of this line of analysis cannot be overestimated, because it indicates the possibility of constructing a unified theoretical framework for the analysis of firms, in which the concepts of contracts, growth processes, capabilities, and historical contingency can be applied simultaneously to the study of business organisation.

***

In the Hayekian conceptualisation of the world (in Moldofsky, 1989, p.101) different mechanisms/models of social and economic coordination are essentially different forms of bringing about order, that is "a state of affairs in which a multiplicity of elements of various kinds are so related to each other that we may learn from our acquaintance with some spatial or temporal part of the whole to form correct expectations concerning the rest, or at least expectations which have a good chance of proving correct"[17].

In the social and economic realm, order manifests itself by a certain correspondence of the expectations concerning the actions of others on which our plans are based with what they will really do. Such state of affairs is the result of two fundamentally distinct sources of order acting in combination: *grown or spontaneous, self-generating order* based on abstract relations of a certain kind; and *concrete order* made deliberately by exogenous forces resting on hierarchical structures of commands and obedience.

Spontaneous orders (of which the market is perhaps the most prominent example) are eminently abstract in the sense that they do not obtrude themselves on our senses (we cannot intuitively perceive them). In order to be aware of their existence we must mentally reconstruct them by tracing the relations established between their parts. Not having been made by an outside agency they cannot be said to serve any specific purpose, even though an awareness of their existence may be

[17] Hayek stresses the fact that this definition draws explicitly on both L. S. Stebbing's (1933, p. 228) assertion that "when we know how a set of elements is ordered, we have a basis for inference" as well as on Immanuel Kant's definition "Ordnung ist die Zusammenfügung nach Regeln" (Order is an integration of regularities,*Werk* (Akademie Ausgabe), *Nachlass*, Vol. 6, p. 669).

critical for their elements to be successful in the pursuit of their own interests. The construction and preservation of such orders is strictly dependent upon the maintenance of a particular structure of relationships between their constitutive elements. This implies they have the potential to reach a level of complexity that far exceeds what could have been achieved through intentional human design. The existence of regularity in the way elements respond to their environment confers a general character to the resulting order, while the detail of its particular manifestation depends additionally on the initial position of those elements as well as on the particular circumstances of the immediate environment to which each element reacts in the process of formation of that order. Therefore, spontaneous orders are adaptations to and incorporate information on a myriad of particular facts dispersed amongst their various parts which will never be known in their totality to anyone. By changing some of the rules of conduct to which the elements obey we may be in a position to influence the general character of the resulting order but never its detail.

In modern societies, the market epitomises a special kind of spontaneous order produced through people acting within the rules of the law of property, tort and contract (Hayek in Moldofsk, 1989). In this game the outcome of each player is determined by a mixture of skill and chance. Individual actions are guided in a way that leads to mutual adjustment of expectations, thorough utilization of the knowledge and skills of all members of society. An efficient state of affairs is thus achieved, in that no need is served at the cost of withdrawing a greater amount of means from the use for other needs than is strictly necessary to satisfy it. Furthermore, and most importantly this game is exclusively means-connected since it makes the reconciliation of divergent purposes possible without requiring any form of agreement on the importance of the needs and associated claims of its participants.

Concrete orders on the other hand are human-constructed entities that can be intuitively perceived by inspection and that are especially designed to deal with directed social order of the kind that bounds organisations. These orders are created by exogenous forces channelled through hierarchical structures of command and obedience. Despite being intrinsically confined to levels of complexity manageable by their creators (whose specific purposes they are intended to serve), man made orders are for a vast number of limited tasks the most powerful method of achieving effective coordination since they enable a degree of control over the particular contents of the resulting order that could not ever be attained through an alternative

ordering mechanism of a spontaneous kind. That is so because the formation of spontaneous orders stems from their elements following rules[18] that are independent of any common purpose in their responses to the immediate environment, while in the case of organisations the emerging order is created through a combination of commands and rules subsidiary to those commands. In typical organisational contexts the general aspects of the upcoming order such as the function to be performed by each element, the purposes to be achieved and the methods to be employed are usually set by commands. However, the details of the procedures are usually governed by subsidiary rules that leave margin for each individual to use his particular knowledge and skills for the performance of assigned tasks (the only way of making use of knowledge spread amongst the different members of the organisation and which nobody possesses as a whole). As Richardson (1998, p.51, 56) puts it:

***

Direction, of the kind that concern us, does indeed mean telling people what to do, but in the sense of allotting them roles and providing associated job descriptions. Management's function is to establish an appropriately related set of roles and rules – an appropriate organisation – and not to seek to prescribe what people should, in every foreseeable set of circumstances, seek to do. Management will change the organisation from time to time as circumstances change and to take account of personalities within a business, but will not routinely intervene to instruct a particular job holder to take this or that particular action at a particular time.

Co-ordination through direction is not, therefore, what might at first seem. It does not mean that those who undertake the activities to be co-ordinated are being continuously told how to carry them out. It implies, rather, the setting up, for a chosen purpose, of an organisation in which people with appropriate skills, aided by appropriate fixed equipment, are given appropriate roles. The roles established, and the relationships between them, are designed to be such that if all the members of the organisation further the particular aims set for them the aim of the organisation will itself be realised.

***

Within a firm, it is often remarked, hierarchy prevails. It does, but one should not imagine that, for this reason, the firm works by means of commands which originate from the top and are filled in with appropriate detail at each operational level. It is doubtful that even armies and navies function

[18] Some of these rules also have a spontaneous character as in the case of those that result from cognitive similitude in the interpretation of the environmental context and behavioural regularities derived from shared cultural patterns, while others are deliberately designed (namely the laws enforced by the State): "some such rules all individuals of society will obey because of the similar manner in which their environment represents itself to their minds. Other they will follow spontaneously because they will be part of their common cultural tradition. But there will be still others which they may have to be made to obey, since, although it would be in the interest of each to disregard them, the overall order on which the success of their actions depends will arise only if these rules are generally followed" (Hayek in Moldofsky 1989, p. 109).

in this manner, far less so-called "command economies". Management does not consist just in giving orders from the bridge though this is part of it, but, rather, in creating, monitoring and, when need be, modifying a system of working relationships designed to ensure that each person employed by the organisation, by doing the specified job allotted to him, will further an overarching purpose. A chief executive can scarcely do more harm than by spending all his time telling people what to do.

***

At the societal level, efficient economic coordination will always rest upon a combination of both spontaneous order and deliberate organisation. At the micro level, coordinating the efforts of limited groups of men with a view to achieving some particular ends will be more adequately fulfilled in the context of organisational entities structured according to hierarchical-authoritarian guidelines. At the macro level, synchronising the activities of these several organisations and separate individuals will be accomplished more effectively through forces conducive to the formation of spontaneous order. Our ability to influence such an order will be confined to some general aspects that define its overall nature and not the particular circumstances through which it will reveal itself in practice. However, that is the way efficient coordination can be brought into a system comprising elements of such number, diversity and variety of conditions that no human mind could ever supervise.

In between the macro and micro level we can think of a third intermediate layer (*meso level*) aggregating forms of inter-organisational coordination premised on a hybrid ordering logic. Hayek does not explicitly acknowledge these structures in his discussion of spontaneous and concrete order. However some of his writtings seem to herald the possibility of such a conceptualisation:

***

"The spontaneous order which we call society also need not have such sharp boundaries as an organisation will usually possess. There will often be a nucleus, or several nuclei, of more closely related individuals occupying a central position in a more loosely connected but more extensive order. Such particular societies within the Great Society may arise as the result of spatial proximity, or some other spatial circumstances which produce closer relations among their members. And different partial societies of this sort will often overlap and every individual may, in addition to being member of the Great Society, be a member of numerous other spontaneous sub-orders or partial societies of this sort as well as of various organisations existing within the comprehensive Great Society" (in Moldofsky, 1989, p.111).

***

2.5 Conclusion

NIE conceptualises governance as an exercise of organisation aimed at minimising transaction costs defined in a narrow sense – costs deriving from combinations of incomplete information, bounded rationality, and opportunistic behaviour. By shedding light over specific aspects of transactions that can generate significant costs of this kind (e.g. asset specificity), NIE provides a useful set of tools with which to understand defining elements of the optimal boundaries of the firm (including public service providers). The approach, however, it is not exempt from limitations.

The study of the internal characteristics of hierarchies and their costs by new institutional economists remains an underdeveloped area. Whether or not the transaction costs apparatus is helpful for shedding light over the fundamental characteristics of the different modes of governance that determine their comparative efficiency in coordinating different kinds of transactions is the subject of controversy. Some authors argue that transaction costs concern exclusively market exchanges, and that the analysis of administrative/bureaucratic costs requires employing other research tools (Demsetz, 1988a; 2002). Other authors lay emphasis on the need to better identify the costs involved in the trade-off among different organisational arrangements (Masten *et al.*, 1991; Joskow, 2005).

The importance ascribed to the vertical integration decision is responsible for the construction of rather narrow and simplistic theory. It avoids a variety of complications that emerge in transactions between firms and consumers, including disparities of information; differential access to technical and legal expertise; differential capacity to bear risk, etc (Williamson, 2004). By doing so, it ignores attributes of transactions other then asset specificity, uncertainty, and frequency that are likely to be critical for describing transactions and that should be factored in.

The next chapter brings together elements of NIE and resource-based theories of the firm under a broader conception of economic organisation. The purpose of such an exercise is to overcome the limitations of both NIE and resource-based theories to explain economic organisation, and to build a more comprehensive understanding of the relationship between governance and efficiency.

Chapter 3: Modes of Governance and Efficiency

3.1 Introduction

This chapter presents the theoretical framework that underpins the research hypotheses on the relationship between governance and efficiency empirically tested in the second part of the thesis.

The expansion of production in modern economies has been accomplished in large measure through specialisation, according to which any one individual performs only a tiny fraction of the kinds of tasks required to make what he or she consumes. The benefits of this process, however, can only be extracted if some degree of coordination between the actions of each individual is achieved. *Coordinating* the actions of the various individual actors so that they form a coherent plan and *motivating* the actors to act in accordance with the plan are the main tasks of economic organisations:

***

"…created entities within and through which people interact to reach individual and collective economic goals. The economic system consists of a network of people and organisations, with lower-level organisations linked together through higher-level organisations.

The highest-level organisation is the economy as a whole. While it is somewhat unusual to think of an entire economy as an organisation, this perspective is useful because it emphasizes that the economic system is a human creation and because many of the problems that smaller, more formal organisations face exist at the economy-wide level as well".

Milgrom and Roberts (1992, p.19)

***

Economic organisation is essentially about bringing order into systems of human interaction. Order can be accomplished through a multitude of modes of governance, a continuum ranging from hierarchies to markets, i.e. from conscious deliberate design to spontaneous self-reinforcing order stemming from the aggregation of self-interested individual actions.

This notion of economic organisation is analytically distinct from that of *formal organisations* (corporations, government agencies, labour unions, universities, churches, etc). The first and foremost defining characteristic of formal organisations is their independent legal identity, enabling them to enter binding contracts and to seek their court enforcement in their own name. This characteristic has been

extensively explored by contractarian theories of organisation[19], which see firms as nexus of contracts, treaties and understandings amongst their members, i.e. legal fictions that allow members to achieve their aims without requiring the operation of complex, multilateral agreements between them. The second defining characteristic of formal organisations is their functional autonomy (they are largely free from intervention by outside parties in their internal operation). This second feature helps to draw the effective boundaries of formal organisations by intrinsically incorporating other aspects of organisational architecture apart from the legal element[20].

Formal organisations tend to be associated with a specific form of economic organisation - the hierarchical mode of governance. This association is far from being univocal and deterministic however. The transactions taking place within formal organisations do not necessarily have to be coordinated by command and authority. The extent to which the underlying governing structure approximates the hierarchical ideal will vary in accordance with the nature of the transactions themselves.

The fundamental unit of analysis in economic organisation theory is the transaction, defined as the transfer of goods and services from one individual to another. Central to economic transactions are individual human beings in the role of indivisible decision makers and actors who create and manage organisations, the performance of which can only be judged with reference to their needs, wants and personal objectives. The key problem of economic organisation - the coordination of individual decisions and actions so that the potential gains of productive specialisation and cooperation may be appropriated - can thus be aptly defined as a problem of optimising the way transactions are conducted (i.e. a problem of organisational efficiency).

Achieving efficiency in the coordination and motivation of the decisions and actions of producers and consumers depends on how the information required for optimal allocation of resources and subsequent adaptations is managed (information about individual tastes, technological opportunities and resource availabilities). The

[19] Originally advanced by Alchian and Demsetz (1972).

[20] Examples of these elements are the patterns of resource and information flows; the authority and control relationships and the distribution of effective power; the allocation of responsibilities and decision rights; organisational routines and decision-making processes; the methods for attracting and retaining members and resources; the means by which new ideas and knowledge are generated and diffused throughout the organisation; the adaptation of the organisation's routines to reflect and implement organisational learning; the organisation's expressed objectives and the strategies and tactics employed; and the means used to unify the goals and behaviour of the individual members of the organisation and the objectives of the organisation as a whole (Milgrom and Roberts, 1992, p. 20).

high degree of dispersion of this information throughout the economy, i.e. the fact that it is not known to any single person or institution in society as it is mainly localised and diffused amongst economic agents, largely contributes to obscuring the process of setting up an optimal mechanism for coordination. In theory two general solutions are feasible. Either transmit the dispersed information to a central planner who is then left with the task of computing the data and providing a solution to the resource-allocation problem, or else put in place a more decentralised system that involves less information transmission, delegating at least part of the decisional power to those who actually hold local information. The main challenge to the first option resides in shaping the system in a way that enables timely decisions to be made while still moderating the costs associated with information communication and planning. The alternative of decentralisation, in turn, needs to tackle the problem of ensuring that autonomous individual decisions drawing extensively on local knowledge yield a coherent and synchronized result.

These two extreme solutions to the problem of economic organisation are epitomised by the traditional, long-debated dichotomy that opposes a system of markets and prices to a system of central planning of economic activity. Taking the market mechanism of coordination to its limit, all transactions could take place between separate individuals on an arm's length basis without there ever being the need for the creation and maintenance of other kinds of organisation except for the market system itself[21]. Conversely, accepting the prescriptions of a system of central planning in its full expression would involve the complete elimination of the price system and the subsequent absorption of all relevant decisions within a single hierarchical organisation. The same line of reasoning can be adopted, without revision, as we move down the scale from coordinating complex systems of economic activity at the national level to simpler and more localised transactional systems like those of small business enterprises. Hence, as two extreme modes of governance, markets and hierarchies are in theory applicable to any problem of economic coordination irrespective of the operational scale.

[21] In such a world each stage of production would be organised as a separate entity/firm and the transactions between stages would be intermediated by the market. If we picture a production line, it is as if each person/firm is buying input from the person/firm on one side and selling output to the next person/firm in the production line. Examples drawn from the private sector, namely comparisons between American and Japanese car manufacturers, seem to demonstrate that this kind of alternative is far from being surreal.

In practice, however, efficient coordination of economic affairs requires the coexistence of both mechanisms. History shows that even in the most centralised economic systems a vast amount of decisions were left to the individual sphere and were partially guided by prices. On the other hand, modern capitalist societies are made of formal organisations interacting with each other through the market, and within these organisations transactions are coordinated through hierarchical structures of authority and command.

The reason why the optimal solution for coordinating economic activities involves a mixture of market and hierarchy is to do with the fact that while these mechanisms can in principle be used in the coordination of any kind of transactions, the efficiency attained by each mechanism will vary with the nature, magnitude and complexity of the coordination process required.

In the presence of problems with *design attributes* (Milgrom and Roberts, 1992), in which there is a great deal of *a priori* information about the form of the optimal solution, i.e. about how the variables should be related, and where failing to achieve the right relationship among the variables is generally extremely costly, centralised coordination tends to reduce the cost of errors and the amount of communication and search necessary to identify the optimal solution. Problems with *innovation attributes*, in which the optimal allocation of resources depends on information not accessible to any of the participating individuals, also require centralised coordination for gathering, developing, and communicating that information to the relevant stakeholders. Finally, problems with public good attributes, externalities, merit good attributes, and/or information asymmetries also lead decentralised coordinating mechanisms to produce inefficient outcomes. In this case, *public* central coordination is required to generate an efficient allocation of resources.

At the other end of the spectrum, it has been formally demonstrated that, under specific conditions, no system can solve the problem of coordinating economic activity at the societal level more efficiently than a system of markets (Arrow and Debreu, 1954). The price mechanism requires the transmission of less information than any other system capable of producing an equally effective outcome. In a perfectly competitive market the transmission of detailed information (about preferences, technological possibilities, resource availabilities, etc) is reduced to a minimum. The knowledge synthesised by the price system ensures that individual

decisions are harmonized while conveying less information than that required by central coordination to achieve a similar outcome. Alongside this information-saving feature, equally valuable is the market's motivational power, that is its ability to channel individually self-interested behaviour towards the generation of an efficient, coordinated collective pattern of choices.

Different modes of governance yield different levels of efficiency depending on the nature and scale of the transactions upon which they are deployed. In other words, different modes of governance entail different costs in coordinating transactions, and these costs should be the decisive criterion to determine which transactions ought to be coordinated through markets and which should instead be centrally coordinated.

Transactions costs comprehend all the costs of running a governance mechanism, including costs of coordinating and motivating individual/organisational decisions.

In a market system, transaction costs of coordination are reflected in the determination of prices and other details of the transactions; in the dissemination and acquisition of information on the existence and location of potential buyers and sellers; and in the process whereby buyers and sellers convene to transact[22]. In a hierarchical system, transaction costs of coordination emerge in the transmission of the dispersed information required to elaborate a central plan; in processing that information so as to devise the plan itself; and in communicating the ensuing guidelines to those responsible for implementing it. During the course of this process there are costs borne out directly from the compilation and transmission of information but also time costs of delay between the moment decisions need to be made and the instant the planned instructions are actually communicated and implemented. Finally, there are also costs of maladaptation resulting from unavoidable imperfections (incompleteness and inaccuracy) in the collection and transmission of information as well as from the limited ability of central planners to process the data received.

[22] In practical terms, on the sellers' side transaction costs of coordination in a market system encapsulate expenditures on such things as market research, advertising and other marketing categories, managerial decisions determining the prices to charge, etc. From the buyers' perspective they condense the costs associated with the time spent searching for suppliers and for the best contractual terms. Furthermore, they include lost benefits not realised due to imperfect matching of buyers and sellers.

Transaction costs associated with the motivational dimension of the organisational problem comprise both costs of informational incompleteness and asymmetries and costs of imperfect commitment. In the first case mutually advantageous transactions fail to occur or costly arrangements need to be put in place on account of lack of relevant information to determine whether the terms of the agreements are mutually acceptable and/or are actually being met. In the second case the incapacity of contractual parties to commit to their own promises and threats creates missed opportunities and/or expenditures intended to facilitate commitment or protect against opportunism. In spite of manifesting themselves in different ways, both types of costs are present in market and non-market governance mechanisms. The diverse nature and intensity they assume under each organisational form implies that for the coordination of specific type of transactions, one form may be better suited than another.

3.2 The Meaningfulness of a Discrete Structural Typology

In the past, some authors have challenged the analytical meaningfulness of the discrete structural typology "market-hybrids-hierarchy". It has been suggested that the study of social and economic coordination requires the employment of different models in combination or in a comparative framework (Goldberg, 1980; Stinchcombe, 1985; Eccles and Crane, 1987; and Bradach and Eccles, 1989). The idea is effectively captured by Thompson's (1991, p.2) analogy of models-torch:

***

"…by shining our model-torch on the complexity of social existence we only expect it to highlight *some* of that complexity. In our case we clearly have three torches to work with – one labelled hierarchy, one labelled markets and the third labelled networks. If we were to shine each of them on to that aspect of running the country under examination, we would expect each torch to highlight just a part of the landscape. Perhaps one or other of the torches would highlight more than the others do. As a result we can say which of the models was most appropriate in analysing that particular aspect of social existence under scrutiny. So our models act as a kind of sifting device".

***

Taken to an extreme, this argument has been used to raise doubts over the analytical value of the distinction between markets, hierarchies and hybrids. Some authors have claimed that organisational reality is exclusively composed of *hybrid*

forms of coordinating exchanges, i.e. interdependent combinations of price, authority and trust in ways that are *not* compatible with a mutually exclusive vision of the relationship between their respective models of coordination:

***

"…it is becoming clear that market, hierarchy and relational contracting are not mutually exclusive control mechanisms. For instance, relational contracts are frequently laced with elements of hierarchy…markets exhibit traits of hierarchies, and hierarchies display properties of markets…These studies suggest that the control mechanisms, price, authority, and trust – which map roughly on to market, hierarchy, and relational contracting – are useful concepts provided we recognize that they are *independent and can be combined with each other in a variety of ways*".

(Bradach and Eccles, 1989, p.99)

***

This argument is founded in a reductionist, linear association of each model to a specific and distinctive *coordinating/control mechanism* (price, authority and trust, respectively). Thompson (1991, p.15) contends that "if it is price competition the central coordinating mechanism of the market and administrative orders that of hierarchy, then it is trust and cooperation that centrally articulates networks". In the same vein, for Bradach and Eccles (1989, p.99) "the control mechanisms, price, authority, and trust map roughly on to market, hierarchy, and relational contracting". Often this linear association between structures and mechanisms is stretched to the point of developing into an *identity*. Hence markets become mere synonyms with prices; hierarchy is subsumed under the label "authority"; and hybrids are equated to trust.

The approach to economic organisation adopted in this thesis deviates from this argument on the grounds that it ignores that each of the three elements (prices, authority and trust) *per se* are not exclusive to, or defining features of each model.

Markets require a certain level of trust in order to work properly[23] as well as coercive authority in the form of law. In order for markets to operate efficiently transactions need to be framed by credible commitments. These commitments are

[23] As both Arrow (1974) and Luhmann (1979) have noted, trust is a remarkably efficient lubricant to economic exchange since in trusting another party one treats as certain those aspects of life which modernity rendered uncertain, reducing complex realities far more quickly and economically that prediction, authority or bargain.

supported by formal mechanisms such as specific contractual clauses, but also informal mechanisms where trust and reputation play an important role.

Hierarchies are not exclusively dependent on authority. As we mentioned in the previous chapter, in order to provide credible alternative to markets as a governance mechanism when tight coordination is required, hierarchies entail a complex combination of control, cooperation, and communication. In particular, they require subsidiary rules that delegate power for individuals to make decisions on the basis of their judgments about the details of each situation (which inevitably involves some degree of trust). In fact, as the level of organisational complexity increases the critical drivers for control shift from enforcement of specific commands to reliance on these rules.

Finally, hybrid forms of organisation do not rely exclusively on trust as a coordinating mechanism either. Hybrids are manifestly unsustainable if deprived of structures of authority for articulating their behaviour and solving possible disputes between their members. They are equally unsustainable if deprived of decentralised coordinating mechanisms for the conduction of operations where local knowledge is of paramount importance.

Central to the theory explored in this chapter is the idea that markets and hierarchies are the two basic, distinct forms of bringing about order in socio-economic systems (Hayek, 1973).

By means of spontaneous ordering forces the market mechanism exploits knowledge dispersed amongst numerous individuals without it ever being concentrated in and limited to the planning of a single mind. This facilitates coordination at a level of complexity[24] that could not have been mastered by deliberate human design. Such capacity however comes at the cost of external inability to control the concrete content of the resulting order.

For the performance of limited tasks (such as those which bring together groups of individuals into organisations), *concrete* order deliberately created through exogenous forces resting on hierarchical structures of command proves to be the most efficient coordinating mechanism insofar as it enables a superior degree of control

[24] Both in terms of the number of the elements involved and the intricacy of the relationships between them.

over the ensuing order, forcing individual action to align with the specific purposes of the ruling authority.

These two key mechanisms of coordinating economic transactions have rules of their own that cannot be adulterated in the way we find most convenient:

***

"Though spontaneous order and organisation will always coexist, it is still not possible to mix these two principles of order in any manner we like. If this is not more generally understood it is due to the fact that for the determination of both kinds of order we have to rely on rules, and that the important differences between the kinds of rules which the two different kinds of order require are generally not recognized".

(Hayek in Moldofsky, 1989, p.112)

***

Amid the ideals of spot market transactions and pure hierarchical structures of command there are hybrid, specially crafted modes of governance (amongst which typical inter-organisational coordination is included) resting on intermediate levels of (de)centralisation of information. Hence, hybrids fall in between the two extremes of fully decentralised organisational systems entirely driven by market forces on the one hand, and fully centralised hierarchical/bureaucratic authority-driven organisations.

The triptych "hierarchy-hybrid-market" contains different approaches to the generation of order. They may not be perfectly descriptive of social and economic systems. However, as ideal references as to how coordination can be brought about they allow us to make progress in understanding the extraordinary diversity of economic arrangements found in empirical reality.

3.3 Decentralised Coordination - Spot Market Transactions

The problem of economic organisation is one of providing individuals with the information required for them to make coherent decisions (i.e. parts of an efficient overall plan), and to motivate them to perform their parts of the plan. A critical claim of neoclassical economics and perhaps the most prominent result of the economic science at large is that, under certain circumstances, a system of prices is able to perform such tasks more efficiently than any other system.

According to the first Fundamental Theorem of Welfare Economics (Pareto, 1906; Barone, 1908; Lerner, 1934; Hotelling, 1938; Lange, 1942; Allais, 1943), if each productive unit knows the prices, its own individual production technology and maximises its own profits at the prevailing prices; if each consumer knows the prices and his/her own individual preferences and then maximises utility given the prevailing prices and his/her income; and if prices are such that supply equals demand for each good and service, then the ensuing allocation of resources is Pareto efficient - given all the restrictions in terms of available resources and technological possibilities, no other allocation is unanimously preferred.

Individuals and organisations acting on a self-interested basis, combining their local knowledge with the information synthesised by prices, and attempting to maximise their utilities and profits are motivated to carry out a plan that will lead them to an efficient outcome. The process whereby the price mechanism leads to a Pareto efficiency hinges on the notion of *rate of substitution*. Leaving the mathematical formalisation of the concept aside, Friedrich Hayek provides an insightful synthesis of the role rates of substitution play in the coordination of transactions, and how prices relate to them:

***

"The crucial point here – which, it must be admitted, even leading classical economists down to John Stuart Mill did not understand – is the universal significance of changing rates of substitution between different commodities. This simple insight, which helped us at last to understand the role of differences and variability of the prices of different commodities, began slowly to develop with the recognition – I will not say the discovery, since of course every simple peasant knew the facts if not their theoretical significance – of decreasing returns from successive applications of labour and capital to land. It was next found to govern, under the name of decreasing marginal utility, the rates of marginal substitution between different consumers' goods. And it was finally discovered to be the universal relation prevailing between all useful resources, determining at once if they are economically the same or different, and if they are scarce or not. Only when it was understood that changing supplies of the different factors of production (or means of satisfaction) determines their variable marginal rate of substitution, was the indispensability of known rates of equivalence (or rates of marginal substitution) for any efficient calculation fully understood. Only when it was at last seen that through market prices this rate of equivalence in all their different uses, mostly known only to a few of the many persons who would like to use them, could be made equal to the rates at which any pair of commodities could be substituted in any of its countless uses, was the indispensable function of prices in a complex economy fully understood.

Variable "marginal rates of substitution" for different commodities, to which I have previously referred, naturally mean their temporary rates of equivalence determined by the situation at the moment, and at which these things must be substitutable at the margin in all their possible uses – if we are to get their full capacity out of them.

It was both the understanding of the function of changing rates of equivalence between physically defined objects as the basis of calculation, and the communication function of prices which combined into a single signal all the information on these circumstances dispersed among large numbers of people, which at last made it fully clear to every person who could follow the argument that rational calculation in a complex economy is possible only in terms of values or prices, and that these values will be adequate guides only if they are the joint efforts, such as the values formed on the market, of all the knowledge of potential suppliers or consumers about their possible uses and availability".

(in Moldovsky, 1989, p.132)

***

A critical aspect of a competitive market system is that it involves the minimal information transmission compatible with the determination of an efficient allocation of resources[25]. A price-guided, decentralised system of coordination does not entail the transmission of detailed information about resource availability, consumer preferences and technological opportunities that a centrally planned system would do. Decisions on resource allocation are left to the individuals and organisations with whom context-specific knowledge on preferences, endowments and production possibilities reside[26]. The price system conveys to the individuals the minimum additional knowledge they need to fit their plans with those of others, and by so

[25] The Hurwicz criterion is one of the most widely applied references for judging informational efficiency: "Suppose that there is no *a priori* information about the optimal resource allocation, so that given what any single producer or consumer knows, any allocation of society's limited resources might still be efficient. Suppose too that each producer is uniquely well informed about its own productive capabilities and each consumer alone knows his or her own preferences and what amounts of various goods he or she initially owns, so that no single agent alone has the information needed to compute an efficient allocation of resources. Then any system capable of supporting an efficient resource allocation using augmented plans must communicate, in addition to the plan, at least one additional variable for each separate good or resource, minus one" (in Milgrom and Roberts, 1992, p.102).
In conditions of perfectly competitive equilibrium the price system communicates one additional variable (the price) for each good after the *numeraire* (the good with reference to which other goods' prices are expressed, historically gold or silver) and achieves economic efficiency with an absolute minimum of information transmission.

[26] The importance of this feature has been vividly captured by Richardson (1998, p. 53): "against the disadvantages of the decentralised approach have to be set the difficulties, which complexity soon make insurmountable, inherent in the timely and continuous collection of, processing of, and acting upon information that centralised deployment would require. Decentralisation has advantages, of course, in terms of motivating and developing those taking the decisions, but recourse to it is essential, most fundamentally, because complex tasks requiring much detailed and "local" information have to be broken down into parts, each of which becomes the responsibility of particular persons or group".

doing, brings a coherent pattern into the economic system - an admirable mechanism given that "it is not the product of human design and that the people guided by it usually do not know why they are made to do what they do" (Hayek, 1945, p.527).

Perhaps one of the most impressive expressions of the power of the market mechanism is the fact that formal hierarchical organisations are increasingly recreating the functioning of the market in conducting their own operations. Modern multidivisional organisations, for example, make extensive use of financial controls, performance measurement indicators and internal transfer pricing systems for coordinating intra-organisational transactions, which represents an obvious attempt to mimic some of the characteristic features of markets as governance devices. The multidivisional form seeks to put in place mechanisms that ensure decisions made at the local level by those with whom the relevant local knowledge resides are coordinated and guided by proper incentives. Senior management usually retains primary roles in critical areas where global coordination is required such as in raising outside capital, allocating resources among divisions, appointing and evaluating divisional managers, and mastering the firm's overall strategic direction. With some degree of variation all other decisions and activities are delegated to divisional managers who normally control context-specific variables such as research and development, design, engineering, procurement, personnel, manufacturing, marketing and sales[27].

3.3.1 Decentralised Coordination and Public Service Provision

Recent reforms in the public sector also provide interesting examples of how some of the virtues of the market are being recreated in areas traditionally coordinated through centrally-driven modes of governance.

In the provision of primary and secondary education services in the OECD for example, although governments are still responsible for most of the funding, they are increasingly relying on the transfer of public funds to privately managed schools as a means of defining an institutional environment that induces efficient behaviour. Furthermore, successive reforms throughout the last decades have transformed the way in which the public sector operates as a service provider. The drive for efficiency

[27] "As Voltaire said of God, if prices did not exist, we should have to invent them, and in the design of the internal arrangements of a firm, this is what we in fact do" (Richardson, 1998, p. 52).

led many countries to opt for decentralising decision-making authority to schools and lower levels of governments. This has been accomplished in the expectation it will enable downsizing of central education administration; elimination of superfluous layers of bureaucracy; and improvement of chains of command in decision making, delivering a larger proportion of financial and human resources directly to local governments, schools, and students (Behrman, *et al.*, 2002). In chapter 5 we find evidence of a negative effect of the share of public providers on efficiency and of a positive effect of decentralisation on efficiency.

In the health sector, increasing disappointment with the rigidity, inertia and apparent inefficiency of bureaucratic national health systems has also triggered a series of market-inspired reforms. In the OECD the quest for efficiency has led many countries to separate purchasers from providers within public integrated systems, to strengthen purchasers' agency role within the health-care system, and more generally to shift towards more independent producers with greater management independence and responsibility. Contracting-out selected activities has also increased, and a limited number of countries have experimented with greater competition among hospitals. Chapter 6 shows evidence of a positive link between some of these reform movements and health care efficiency – public integrated systems *with* active purchaser provider split are found to outperform public integrated systems *without* active purchaser provider split in terms of efficiency.

In the social protection field, the financial sustainability of traditional Bismarckian systems rooted in the state's monopolistic provision of social insurance has also been called into question in recent years. In attempting to protect individuals of certain corporatist groups from inappropriate voluntary self-protection, these systems often fail to protect those who are in most need. They also leave a very small margin for individual voluntary action in the insurance market, which results in a substantial share of social transfers being allocated on a categorical basis and redistributed horizontally. Beveridgian systems, on the other hand, lay emphasis on a basic role for the state in the provision of social insurance. They acknowledge the need for state intervention (generally in the form of flat-rate insurance benefits covering as large a proportion of the population as possible), but leave protection above a given minimum to be acquired in the market through voluntary social insurance. Chapter 7 shows evidence of this market-inspired policy approach clearly

improving the targeting of transfers towards the poor, and therefore the efficiency of social expenditure.

3.4 Centralised Coordination - Hierarchy

One of the basic economic reasons for the existence of the public sector is the correction of situations where the market fails to organise efficiently transactions associated with particular goods, services or activities. When markets fail, however, governments are not the only ones that can step in to remove, avoid or mitigate whatever factors are preventing markets from achieving efficiency. Private individuals and firms also dispose of and in fact exercise the option to use nonmarket forms of organisation that enable them to meet more thoroughly their economic goals (Chandler, 1977).

Formal organisations are one particular expression of such exercise. They represent conscious efforts to supplant the market mechanism when planned central coordination proves to be more efficient a mechanism for coordinating economic activities.

Formal organisations do not always set their dealings with one another on the basis of simple market transactions. They often coexist in dense networks of co-operation and affiliation with other organisations, which represent another example of deliberately created nonmarket forms of organisation. These networks may be geared towards creating stable trading relationships so that demand expectations are made more reliable and thereby production planning facilitated; they may have the purpose of inducing sub-contractors to assume the risks inherent to investments in specialised skills and equipments and permitting continuing cooperation between parties involved in the development of specifications, processes and designs; or they may be a vehicle for pooling or transferring technology, just to name a few examples.

As we have been arguing, the particular governance mechanism that proves to be more efficient in generating economic coordination is a function of the nature and scale of the transactions onto which the mechanism is applied.

Some transactions have to be put into a precise relationship in respect of their character, magnitude and timing as they exhibit properties of *systematic close complementarity* (Richardson, 1998). Milgrom and Roberts (1992, p.91) refer to these transactions as involving *design attributes*, situations in which “there is a great deal of

a priori information about the form of the optimal solution, that is, about how the variables should be related, and failing to achieve the right relationship among the variables is generally more costly then are other kinds of errors, including slight misspecifications of the overall pattern, as long as the individual pieces fit"[28].

Examples of problems with design attributes include tasks of synchronization and assignment. Common to both tasks is a sense of urgency about the decision; the extreme dependence of the optimal course of action on particular circumstances; and the substantial knowledge hold by some central entity about the form of an optimal decision. Under these conditions, using a centralised governance mechanism that directly communicates the design variables themselves to ensure coordination is generally more efficient than a decentralised alternative solution such as the one projected by a system of prices. The inadequacy of the price mechanism in this particular instance lies in the excessive level of information it requires to provide a solution to the coordination problem and also in its difficulty to avoid costly failures of synchronization or fit[29].

When design problems arise repeatedly, calling essentially for similar solutions, centralised coordination is normally abandoned in favour of the implementation of *routines* that ensure decentralised solutions to the recurrent design problems. Routines provide the human assets of formal organisations with the opportunity to exercise judgement and act accordingly, within the general framework set by the subsidiary rules of the organisation. This freedom of action means that the interactions among members of the organisation will tend to establish ever-new rules for the coordination of their activities, although they will always be framed within the rules set by the commanding authority (Ioannides, 2003; Nelson and Winter, 1982). Given their spontaneous nature, these routines can be seen as embodying *tacit* knowledge shared by the human assets of the organisation, enabling them to cope with ignorance and to bring order to their collective actions. In the words of Richard Langlois (1992, p.176), "the personnel of a firm follow, invent, learn and imitate routines that persist over time. As in Hayek's theory of culture, the routines are often

[28] Milgrom and Roberts (1992) also refer to these transactions as exhibiting *design connectedness* properties – transactions extremely sensitive to problems of maladaptation to the system they are part of, which usually leads to the strengthening of central coordination mechanisms.

[29] If some of the information is missing or inaccurate the performance of the price system deteriorates more sharply than in the case of central coordination (the brittleness/sensitivity of the price system is thus greater in comparison with more centralised alternatives).

tacit and skill-like, followed unconsciously because they produced success in the past".

Transactions with *innovation attributes* are another typical domain for central coordination. When the optimal allocation of resources depends on information not accessible to anyone alone at the operating level of the organisation, it becomes necessary for someone to gather, develop, and communicate that information to decision makers in the organisation. Local knowledge may still have to be combined with new information so that effective coordination may be achieved. However, in such circumstances a decentralised system like the price mechanism, built on individual responses formulated on the basis of localised information will often be inappropriate for the achievement of an optimal coordinating plan.

Some of the most important problems of strategic decision-making in formal organisations display both design attributes and/or innovation attributes, which run counter to using prices or other decentralised models of coordination and favour centralised control.

The scale, scope and core competencies of the organisation are all design variables that can be more efficiently administered through a centrally-driven approach. Operational scale is a design variable because it has predictable implications for the various parts of the organisation, and errors deriving from incorrect perceptions of scale by some parts of the organisation can be very costly. In a typical formal organisation the anticipated volume of activity influences any decision on production capacity, size of sales force, supplies, distribution equipment and facilities, etc. Everyone in the organisation needs a shared vision of the intended scale of operations if their actions are to be coherent. Economies of scale achieved in the manufacture of components used in several products[30] also increase the need for central coordination since the optimal scale of each product is an increasing function of the predicted scale of the other products that use the same components. Equally important a kind of economy of scale, entailing similar needs of coordination, is the one that concerns the core competencies of the organisation, the acquisition of knowledge, experience and skills required to design and trade new products in a set of related markets or in using a set of related technologies[31].

[30] Also designated as *economies of scope*.

[31] In a certain sense, these competencies or *capabilities* (Richardson, 1972) are shared components whose costs are spread among a series of products that do not yet exist.

The existence of complementarities between a set of activities is one of the most important sources of design attributes[32]. Organisational strategies themselves are made of complementary dimensions. In an industrial context this is reflected in the fact that a company's business strategy encompasses decisions over essentially complementary issues, including product strategy, manufacturing policy, equipment choice, human resource management, supplier relations, accounting methods, etc. Again, central coordination is required so that the various parts of the organisational strategy may be connected and aligned properly in an optimal way.

In addition to design attributes, organisational strategies also involve innovation attributes since the information necessary to identify the optimal strategy and the specific pattern that enables its pieces to fit together is often not available within the organisation, and thus needs to be acquired and compiled from the outside. Radical changes in the environment (for instance in demand conditions or production technologies), may call for equally radical shifts in strategy. Being confined to the use of local knowledge, the members of the organisation may only engage in processes whereby their decisions are marginally adjusted to the local movements in the environment. However, recognising that radical strategic shifts are necessary and estimating the benefits of such changes is most likely to require knowledge they do not possess and that can only be obtained and efficiently administered through central coordination.

***

A firm is a system of organized co-operation. Where only a few complementary activities have to be fitted together, this can be achieved by co-operation between the relevant arties, each of them having a more or less equivalent say. It is possible for a handful of people to work out together a concerted plan of action and then to be jointly responsible for its execution. But even then it will prove at best laborious to agree whatever modifications in the plan will from time to time be required in order to meet changing circumstances. A leader is likely to emerge, or be chosen, and members of the group will tend usually to follow that leader, without having to be convinced at each and every stage that he or she is on the right road. Where the number of closely complementary activities or investments is large, where they have to be put into a precise relationship in respect of their character, magnitudes and timing, I have said, for want of a better term, that *systematic close complementarity* exists. In such circumstances, co-ordination requires *direction.*

[32] In this context a group of activities is said to be mutually complementary if doing more of any one particular activity increases (or at least does not decreases) the marginal profitability of each other activity in the group. Complementarities introduce predictability into the relationships amongst the activities, which is precisely one of the defining features of design attributes.

***

I cannot therefore agree with Coase that the existence of firms as to be attributed to the cost of using the price mechanism. As we shall see, a cost of this kind may be a consideration, even if not the principal one, in a firm's decision to "make or buy" a necessary input, but to say that transactions costs explain the *existence* of firms is to reduce the role of conscious business planning to vanishing point. The term "price mechanism" may perhaps too readily lead us to think that resources are somehow allocated without human agency. The reality, as we know, is that things happen because people do things; in this context, firms and consumers do things on the basis of plans that they consciously make, but plans which by interacting in their execution are so modified as to become woven into an overarching economic order.

Where ordering can be built up through incremental adjustment it can come about spontaneously, as a result, that is to say, of decisions taken by firms independently, without any consideration being given to their necessary integration. But where ordering relates to a set of closely complementary activities that must fit together in a specific way there must be conscious design and planned execution. And it is the function of management, at different levels within the organisation, to provide this.

George Richardson (1998, p. 49,50)

***

3.4.1 Public Central Coordination

Decentralised coordination of transactions with design and/or innovation attributes is sub-optimal due to the brittleness of the system in dealing with imperfect information and excessive demands in terms of communication. The quest for efficiency then leads these types of transactions to be coordinated through formal organisations structured along hierarchical lines.

Decentralised coordination of transactions with public good attributes, externalities, merit good attributes, and/or information asymmetries is also sub-optimal and leads perfectly competitive markets to produce inefficient outcomes[33]. Under these circumstances, *public* central coordination is required to deal with these particular features of transactions and to generate an efficient allocation of resources.

[33] The effect of increasing returns to scale in some industries is another classic example of a situation where the market, left to its own fate, is due to fail. In conditions of *natural monopoly* Pareto optimality requires a pricing policy that will result in significant losses for producers. From a private point of view the viability of production hinges on cooperation between producers with a view to acquiring market power. From a social point of view, however, such strategy would bring about allocative deadweight losses as a result of the ensuing reduction of output. In these circumstances public central coordination in the form of government regulation, subsidisation of producers or nationalised corporations may prove to be beneficial.

Transactions with public good attributes are non-rival in their consumption, and non-excludable[34] (Samulelson, 1954). Each of these two attributes (or a combination of both) is a potential cause for market failure (Stiglitz, 2000). On the one hand, problems of demand revelation associated with the dominance of non-cooperative strategies enable individuals to "free-ride" by consuming the good without contributing to their provision. On the other hand, it may be inefficient for society as a whole to exclude those individuals from consumption (Head, 1974). The achievement of a cooperative solution in this case hinges on public central coordination and its ability to coerce tax contributions, especially when a high number of people are involved (Olson, 1993).

Transactions subject to externalities reflect situations where the utility of some individuals is influenced by activities of others in a way that is not captured by the price mechanism (Buchanan and Stubblebine, 1962), thereby originating discrepancies between private and social costs and benefits. Public central coordination is required to take account of these external effects, and to achieve an efficient allocation of resources.

The efficiency of decentralised coordination rests on the assumption that individuals are the best judges of their own welfare. Accepting that premise is equivalent to accepting that individuals think they ought to be sovereign consumers, also that they think they are able to be sovereign consumers and finally that they want to be sovereign consumers (Mooney, 1979). While these judgements may be valid in the vast majority of cases, there are situations where they may not hold due to limited information, limited consumer rationality (e.g. mental illness or other personal damage) and/or consumer delegation. In such cases of transactions with merit good attributes, "public policy [should] aim at an allocation of resources that deviates from that which is reflected by consumer sovereignty"[35] (Musgrave, 1959, p. 9) – again for that purpose public central coordination is required.

Deviation from the neoclassical assumption of perfect information is one of the strongest criticisms of a system of freely competitive markets (Arrow and Lind, 1970). When one party to a transaction has more or better information than the other party decentralised coordination is bound to create inefficiency. The effects of

[34] Once the good has been provided consumers cannot (at less than prohibitive cost) be excluded from consumption benefits,
[35] Musgrave classifies these situations as involving *merit goods*.

asymmetrical information are particularly acute in private insurance markets where they give rise to adverse selection and moral hazard problems. In those circumstances, public central coordination is often required so that a pooling equilibrium is reached by making the insurance compulsorily and regulating the ability of individuals and organisations to accept/reject contracts.

3.4.2 Government Failure

The case for public central coordination needs to be balanced with the idiosyncratic limitations of public bureaucracies and the political process. An imperfect market must be compared with an imperfect government. The relevant choice is not so much between an ideal norm (pure and perfect competition) and an existing imperfect institutional arrangement (the "real" market), but rather between alternative real and imperfect institutional arrangements (the market with or without external corrections)[36].

Numerous characteristics specific to political bureaucracies hinder the efficiency of governmental coordination. These need to be taken into account when calls for public central coordination are being made (Stiglitz, 2000).

The *organisational incentives* of public bureaus are normally driven by political concerns and provide little incentive to maximise productivity. They are often monopolies operating in "non-contestable" markets, i.e. protected from potential competition due to possible entrance of other firms in the market. Because of public ownership, these monopolies face soft budget constraints (any losses are made up out of government revenues, and there is no threat of bankruptcy – Komai, 1986).

The organisation of the public sector obeys *civil service rules* designed to ensure that public employees do not abuse their position and power, that governments hire and promote people on the basis of merit, and that their pay is appropriate. While these rules serve an important function, they introduce substantial rigidities to the management of personnel with significant effects in terms of incentives. In a similar vein, traditional public bureaucracies are subject to numerous *procurement restrictions*, destined to avoid abuses in government's purchases, but with similar consequences on incentives and efficiency. The *political restrictions on budget-*

[36] An ideal norm can only be used as a reference or standard from which divergences may be assessed for all practical alternatives and the most efficient alternative selected (Demsetz, 1988c).

setting processes are another source of inefficiency, particularly strong for long term capital investments.

In the absence of profit-driven incentives and performance related rewards, individuals are often encouraged to pursue political goals, such as the power and social prestige of supervising large organisations. Rational, self-interested bureaucrats will tend to expand the production of public services beyond the socially optimal level (Niskanen, 1971). The development of such *bureaucratic behaviour* detrimental to efficiency is stimulated not only by improper individual incentives but also by the general lack of competition.

One trait typical of public bureaucrats is their excessive levels of *risk aversion*. By following certain bureaucratic procedures that ensure that all actions are peer reviewed, public bureaucrats are able to absolve themselves of responsibility for errors. Since they do not bear the costs of excessive focus on bureaucratic procedures (the ensuing inefficiency is supported by the general body of tax payers), there is a tendency for these formal procedures to thrive[37].

Public choice theorists have also emphasised that public organisations are prone to failure because of three additional factors. First, there is no credible reason to believe that politicians will demand a pattern of goods and services that reflects the interest of society as a whole. Politicians act in accordance with their own interest, on many occasions producing benefits for interest groups[38] as a result of vote trading, or pursuing policies involving short run highly visible outputs in detriment of the more beneficial, but less visible, medium and long run projects (Mueller, 2003). Moreover, the nature of the political process itself encourages imprecision in policy formulation. Clearly defined policies allow failure to be easily spotted. Yet, lack of precision in policy formulation is in itself a source of failure. Second, there is also no convincing reason to believe that the bureaucracy itself will carry out the instructions of politicians, since these instructions may well not be in line with the bureaucrats' interests. Thirdly, public organisations are apt to be dynamically inefficient in developing technological and methodological innovations due to lack of competition and inability of individuals to collect the benefits of technical advance.

[37] The need to follow a set of routines is not merely the result of bureaucrats' self-interested behaviour but follows from the fiduciary relationship between government bureaucrats and funds they allocate, i.e. the "publicness" of the later.

[38] Interest groups in turn tend to concentrate on issues related with redistribution of income which yield greater and more immediate returns than pressuring for total production improvements (Olson, 1982).

In general, the mechanisms that bring together supply and demand of public services in public bureaus are weak because it is "essentially a political process characterised by lags, bottlenecks, coalitions, log-rolling and other fuzzy attributes of political behaviour" (Wolf, 1988, p. 62).

3.4.3 Public Central Coordination and Public Service Provision

The virtues of central coordination have long had a strong echo in the provision of public services. In primary and secondary education, for example, governments are often forced to intervene in the funding of the services so as to correct/mitigate the effects of externalities and credit constraints that decentralised coordination fails to address. Externalities in education tend to occur because individuals interacting through a decentralised decision-making process are not likely to take into account the full effects of education on economic growth (Barro, 2001). Credit constraints arise when individuals are not able to borrow against their human capital. This limits the capacity of families to fund school education and to provide an appropriate educational environment during children formative years (Carneiro and Heckman, 2002)[39]. In both instances, governments may intervene by providing financial support and pushing the consumption of education closer to the societal optimal.

The virtues of central decision-making in education appear to extend further to the service provision level. Since the early eighties, educational reforms in OECD countries have shifted decision-making authority in some areas (e.g. organisation of instruction[40]) to lower levels of the education system. However, increased school decision-making autonomy in some areas has been accompanied by greater influence of central authorities in other areas such as planning and structures[41], through national assessment programmes and centrally established frameworks (OECD, 2004a).

[39] Carneiro and Heckman (2002, p.3) refer to this credit constraint as "the inability of the child to buy the parental environment and genes that form the cognitive and noncognitive abilities required for success in school".

[40] Student admissions; student careers; instruction time; choice of textbooks; grouping students; additional support for students; teaching methods; and regular day-to-day student assessment.

[41] Opening or closure of schools; creation or abolition of a grade level; design of programmes of study; selection of programmes of study taught in a particular school; choice of range of subjects taught in a particular school; definition of course content; setting of qualifying examinations for a certificate or diploma; and credentialing (examination content, marking and administration).

In healthcare, decentralised market interaction between patients and healthcare providers exposes individuals to the financial risks of illness unless the financial resources are adequately pooled. The ability of private insurance markets to adequately pool financial risks and promote access to services is hampered by problems of adverse selection, and moral hazard resulting from information asymmetries (Docteur and Oxley, 2003). Higher propensity of those with greater health risks to buy insurance and to insure at higher levels than those in good health limits access to affordable insurance for high risk individuals, lowers coverage and leads to under-consumption of healthcare from a social point of view. On the other hand, the fact that insured households do not bear the full cost of treatment received, leads to consumption beyond the social optimum. Finally, healthcare providers are better informed than insurers about the need and scope for medical treatment, and about the quality of services supplied, which allows them to artificially induce demand for care.

In medical care, a combination of information asymmetries between patients and physicians as to the complexities of medical care diagnosis and treatment, difficulties in ascertaining the quality of services even after they have been provided, and the profit-maximisation motif of private providers disseminate perverse incentives that lead competitive markets to failure.

Public central coordination in the form of both public provision of insurance and public regulation of health service provision and private health-insurance markets is then critical for the efficiency of health systems. This explains why in most cases, the public sector has come to take a dominant role in the financing and, in some cases, the provision of healthcare services.

A similar problematic permeates the field of social protection. In a world of perfect knowledge and perfect foresight, with well functioning capital and insurance markets, there would be no reason for the state to play an extensive role in social protection. Voluntary decisions, insurance contracts, and the saving instruments created by markets would suffice. Government intervention would only be required at a residual level for the provision of social safety nets to those whose total income over the entire life-cycle is insufficient to sustain an adequate living standard. However, real world social insurance markets also suffer from moral hazard and adverse selection problems which prevent them from delivering an efficient outcome. Public central coordination is then required to make insurance compulsorily, and to regulate

individuals' ability to accept/reject so that a pooling equilibrium may be reached. Additional factors that limit the efficiency of social insurance markets include uncertainty underlying the size of future claims (moulded by macroeconomic instability and inflation in particular), and the consequent difficulty/impossibility of setting a realistic price for the insurance; and time correlation of social risks, which determines that the aggregate risk cannot be solved by pooling the individual risks (public central coordination by itself does not solve this problem, but the state may have an edge in using its budget and public debt to shift resources across time). At a more general level, there is evidence that, left to their own decisional criteria, individuals tend to be sub-optimally provident, in that their saving pattern fails to fully acknowledge their future income needs. Public central coordination is then required to overcome individuals' short-sightedness in making provisions for adverse social circumstances in later stages of the life cycle.

3.5 Hybrids

In seeking to constantly improve organisational efficiency, individuals have kept forging innovative hybrid governing structures that combine features of both markets and hierarchies in an attempt to bring the best of the two worlds together. Milgrom and Roberts (1992) present four examples in the private sector that give a good account of such an effort.

- In the nineteenth century, the development of utopian experimental communities led to the creation of the co-op (or cooperative) form of organisation. These business organisations were owned by the individuals who transacted with them (workers, costumers, or occasionally suppliers) and structured around specific legal rules. Individual members´ voting power was usually exerted on a one-member/one-vote basis, contrasting with the traditional rule in standard private sector organisations based on the capital share hold by each member. Profits were distributed in the form of reduced prices for the services or goods supplied. Despite their initial development being formulated on an explicitly social and ideological basis, the popularity of the cooperative form of organisation grew over time as a result of its efficiency in alleviating problems of monopoly pricing in situations where economies of scale restrict the number of suppliers that can profitably operate in the market.

- In the first half of the twentieth century, another example of a hybrid mode of governance appeared in the form of franchise retailing[42]. This was a clear attempt to combine the advantages of ordinary market arrangements (owner-operators' incentives to maximise profits) with the economies of scale arising from selective central coordination (in specific fields such as marketing, purchasing, etc), while simultaneously avoiding the problems posed by asset specificity in spot market transactions. As the parties are unable to accurately foresee the circumstances bounding their relationship in the future, they agree on processes through which decisions are to be made by contractually binding themselves to explicit rights and responsibilities. At the same time they develop routines in the conduct of their business so as to bring an extra degree of regularity to their relationship. The nature of the franchise agreement, as a mode of governance, is thus both different from the pure hierarchical structure of most formal organisations, and the cash-for-product contract of simple market transactions.

- In the second half of the twentieth century, another success story of hybrid coordination of transactions was the supplier organisation introduced by Japanese automobile companies. Instead of relying on vertical integration or competitive bidding as many other systems did (especially in the United States), the Japanese management of supplier relations focused on achieving many of the advantages of market incentives by implementing a reward system whereby suppliers were evaluated in accordance to their performance in previous contracts. In comparison with the traditional method of competitive bidding among independent suppliers, this system enabled companies to assess supplier performance with much greater accuracy. It allowed them to select suppliers before the parts specifications were made final, opening up an opportunity to exploit considerable economies resulting from involving the suppliers' know-how in design engineering; conceiving parts that fitted the capabilities of the suppliers' existing equipment; and giving them more time to plan and prepare for production. Both the Japanese system and the system of vertical integration used in the United States allowed for the creation and protection of specific assets, the former hinging on reputation and repeated dealings and the later on ownership. However, an important advantage of the Japanese system was that it did

[42] The franchisee owns and runs a retail business using the franchisor's brand name, often buying inputs or goods for resale from the franchisor. The franchisor collects fees and royalties from the franchisee for the use of the brand name and commonly also provides training, advertising, and other services. Franchisors also generally maintain rights to set and enforce standards on the franchisee.

not seem to be subjected to the kind of influence costs[43] found in its American counterpart in case of failure. Breaking relations with a supplier who failed to perform to adequate standards often proved to be substantially easier than stopping obtaining supplies from badly performing internal divisions. Furthermore, the Japanese system generated a competitive framework that brought an efficient discipline to prices and quality, in a way that seemed difficult to mimic under common ownership through competing internal supply divisions, also for influence cost reasons.

- Finally, the multidivisional form of organisation mentioned in section 3.3 is in itself a hybrid governance mechanism that does not fit either with the extreme model of spot market transactions or with the pure hierarchical approach to economic organisation[44]. It represents an attempt to tackle problems encountered in both highly functionally centralised organisations as well as in collections of entirely decentralised firms under common ownership (the so-called holding companies). In organisations relying mainly on central coordination the distance separating head-office decision makers from critical local knowledge often poses very significant obstacles to efficient decision-making. Costly time and information lost in communication throughout the bureaucratic structure hampers an effective course of operations, as does the excessive level of complexity managers are forced to contend with. In holding companies, on the other hand, the head office typically performs a very limited or no management role at all, self-confining its actions to collecting the profits of the constituents firms. This system, however, fails when there are gains to be extracted from exploring economies of scale and coordinating decisions across units on production, investment and marketing.

As we have seen in the last chapter, there is a growing body of literature on these assorted organisational forms, highlighting their role in facilitating the stabilisation of relations between independent exchange partners through informal

[43] The result of attempts to reallocate and protect rents and quasi-rents within organisations. In general, the larger the organisation, the larger are the rents in total and the more people competing for them. Bigger prizes and greater number of contestants, tend to create greater competition for these rents, leading larger firms to suffer from more than proportionally higher influence costs (Milgrom and Roberts, 1992).

[44] The way in which information is made use of within multidivisional organisations is set at an intermediate level of (de)centralisation, in between the two limits of individual decisions based on local knowledge complemented by the price system, and centrally-planned coordination resting on a bureaucratic structure of authority and command. In these organisational arrangements, centralised decisions define parameters that constrain the discretionary power of local decision makers.

mechanisms based on trust. The conceptualisation of these patterns of reciprocity and collaboration as alternative governance mechanisms distinct from markets and hierarchies is said to enable firms to gain access to know-how unavailable in-house; spread the risks associated with uncertain ventures; benefit from economies of scale; enter new product and geographic markets more efficiently; manage interorganisational dependencies; and respond adequately to changing circumstances (Pfeffer and Nowak, 1976; Porter and Fuller, 1986; Miles and Snow, 1986; Powell, 1987; Johnston and Lawrence, 1988; Ouchi and Bolton, 1988).

The approach adopted in this thesis takes these hybrid arrangements as resting on intermediate levels of (de)centralisation of information, in between the two extremes defined by fully decentralised organisational systems entirely driven by market forces on the one hand, and fully centralised hierarchical/bureaucratic authority-driven organisations. Hybrids offer the possibility to coordinated transactions through long-term relationships between autonomous organisations. Mediating these relationships is an element of trust that plays an important role for the sustainability of the overall structure. Trust is here taken as the reflection of an enlightened mutual interest (Doz and Hamel, 1998) which leads organisations to renounce the exploitation of profitable opportunistic behaviours in the sort run, for the sake of medium and long term dynamic benefits that outweigh those that would arise from pursuing a static optimisation strategy.

***

The essence of co-operative arrangements such as those we have reviewed would seem to be the fact that the parties to them accept some degree of obligation – and therefore give some degree of assurance – with respect to their future conduct…Where buyer and seller accept no obligation with respect to their future conduct, however loose and implicit the obligation might be, then co-operation does not take place and we can refer to a pure market transaction. Here there is no continuing association, no give and take, but an isolated act of purchase and sale such, for example, as takes place on an organised market of financial securities. The pure market transaction is therefore a limiting case, the ingredient of co-operation being very commonly present, in some degree, in the relationship between buyer and seller. Thus although I shall have occasion to refer to co-operation and market transactions as distinct and alternative modes of coordinating economic activity, we must not imagine that reality exhibits a sharp line of distinction; what confronts us is a continuum passing from transactions, such as those on organised commodity markets, where the co-operative element is minimal, through intermediate areas in which there are linkages of traditional connection and goodwill,

and finally to those complex and inter-locking clusters, groups and alliances which represent co-operation fully and formally developed". (Richardson, 1972, p.886, 887)

***

3.5.1 Hybrids and Public Service Provision

Hybrid modes of governance are not strange to the provision of public services. Most public service governance arrangements observed in practice are hybrid in the sense that they that do not fit entirely in the centralised hierarchical ideal or in the pure spot market transactions model. Furthermore, some of the most interesting examples of alliance/network forms of organisation are being implemented in the public sector.

In the provision of primary and secondary education, the public sector is redefining the way it operates as a service provider, from a traditional purely bureaucratic form of organisation to the constitution of effective service agencies, highly decentralised, and operating along privately-inspired management criteria. In this new environment, partnerships and other forms of collaborative work play a prominent role. Governments are increasingly relying on schools that are privately managed but predominantly financed through the public purse to achieve efficiency gains, simultaneously combining elements of central coordination with decentralised managerial capacity. That same combination is captured in a recent survey on the distribution of the decision-making process in lower secondary education (OECD, 2004a). According to this survey, decisions on the organisation of instruction are predominantly taken by schools in all OECD countries, while decisions on planning and structures are mostly the domain of more centralised tiers of government[45].

In healthcare examples of inter-organisational cooperation also appear to abound. In theory, health services can be organised in three fundamentally different ways: via hierarchical bureaucracies; through long-term contractual arrangements under some degree of non-market control; and as direct, short-term market-based interactions between patients and providers (WHO, 2000). Given the disadvantages of both hierarchical and market-based modes of governing health services, recently many countries have opted for experimenting with long-term contracts to achieve the

[45] The distribution of decision-making authority is less clear for items such as personnel management and allocation and use of resources.

combined advantages of greater flexibility and scope for innovation while maintaining overall control over strategic objectives and financial protection. This has involved transformations to the incentive environment of organisations, modifications to the distribution of decision-making control, revenue rights, and financial risk of the different stakeholders. Chapter 6 shows evidence of these changes having a positive effect on efficiency, and of further movement towards a market-inspired mode of governance having a negative effect on efficiency[46].

3.6 Government and Economic Performance

The effects of public policies on economic productivity come about through two different channels - via productivity changes in the public sector itself, and via the effects of taxation, public spending and regulations on the private sector (European Commission, 2004).

The public sector is responsible for a large proportion of the overall output of an economy. Public employment ranges from 10% (Germany) to 30% (Sweden) of the labour market in the European Union. Public sector labour productivity is therefore an important driver of average labour productivity nationwide.

Furthermore, the organisation and functioning of the public sector affects productivity in the private sector. Taxes distort relative prices in the economy influencing economic incentives, namely the willingness to work and invest. Government expenditure on selective areas such as education, health, research and development or infrastructure is an important determinant of long term economic growth. Government regulations also have an important impact on economic performance by limiting the choices of individuals and enterprises.

There is a growing body of empirical literature investigating the link between government activity and the performance of economic systems via this second channel[47]. The literature stresses the importance of some *core*, *essential*, *productive* expenditure for economic growth. This type of expenditure is said to be as important to growth as private capital and labour. It can act directly upon the human and

[46] According to the categorisation of health systems employed in chapter 6, public contract systems and private insurance/private provider systems are shown to be systematically outperformed by public integrated systems.

[47] Hemming *et. al* (2002); Romero de Avila and Strauch (2003); Tanzi and Shuknecht (2000, 2003); Tanzi and Zee (2000); and Zagler and Durnecker (2003).

physical capital stock and technical progress of the economy, as well as indirectly through synergies with private activities[48].

The first and foremost government-spending category is on administrative services and justice. The creation of a well-defined institutional framework conducive to growth implies the formulation of clear property rights so as to minimise institutional uncertainty, enhance the predictability of returns of investments and thus increase market efficiency. It also implies the enforcement of general rules that promote competition, allow information to flow without restraint and foster efficient risk management. Such an institutional framework will work to minimise transaction costs of economic agents, and help them deal more effectively with externalities and spillovers[49].

Another area where government intervention is usually associated with positive growth effects is investment expenditure[50] (acquisition or accumulation of physical capital goods). Physical infrastructure can be seen as a public good since users usually cannot be excluded from consumption, and one user's consumption might not interfere with the consumption of others. On certain occasions, exclusion might be technically feasible and there may be some degree of rivalry in consumption, however the state may still decide to provide the good/service free of charge at the point of delivery because exclusion is economically inefficient[51] or inequitable. Notwithstanding a tradition of dispute[52], there is currently a consensus over the growth-enhancing properties of public investment. Public expenditure on physical capital is thought to be linked to higher growth rates if it is applied on infrastructures that serve as inputs for the private sector (e.g. transport, communication and public utilities)[53]. There is also evidence of this type of expenditure being exposed to diminishing marginal returns (De la Fuente, 1997; Kalyvitis and Kalaitzidakis, 2002).

[48] Afonso, Ebert, Schuknecht and Thone (2005) provide a good review of this literature.
[49] Van Rijckeghem and Weder (2002); von Hagen, Hallerberg and Strauch (2004).
[50] Aschauer (1989); Cashin (1995) ; Nourzad and Vrieze (1995); Sanchez-Robles (1998); Shioji (2001); and Kamps (2004).
[51] In situations of zero marginal cost from additional users.
[52] Some authors present evidence of a non significant impact of public investment on growth - Barth and Bradley (1988); Ford and Poret (1991); Holtz-Eakin (1994); Yi and Kocherlakota (1996); and Cassou and Lansing (1999). Vanhoudt, Mathä and Smid (2000) even find evidence of reverse causality (economic growth exerting a significant impact on public investment, and not the other way around).
[53] Easterly and Rebelo (1993); Devarajan, Swaroop and Zou (1996); Thöne (2004).

The production of human capital is the area of government activity for which there is the clearest evidence of positive growth effects[54]. Government intervention in education hinges on four main pillars (Hanushek, 2002): externalities, economies of scale, other market failures, and redistribution motives. The available evidence indicates that public spending can push education closer to the optimum in comparison with a pure market scenario. This result is only weakened by the limitation of most empirical studies to formal school education (vocational training is typically kept out of the analysis), and by regressing economic performance on school attendance rates, schooling years, or graduation rates instead of public education expenditure.

Health expenditure is also typically seen as an area where public intervention leads to better economic performance. Government intervention is justified both by equity and efficiency considerations. Health insurance markets are prone to failures due to problems created by information asymmetries. These failures have a negative impact on the efficiency of the sector. Perverse incentives may be created, encouraging over/under consumption/provision of health services. Public sector provision and regulation of health services can mitigate some of these failures, help improve general health condition of the population thus creating a direct positive impact on human capital and hence on growth[55].

Research and development is commonly recognised as a key growth factor in modern economies. Difficulty in confining the commercial exploitation of an innovation to the creative firm is a source of externality failure, resulting in R&D activity being conducted below the social optimum. Government corrective intervention is therefore justified, both in the form of research conducted in public institutions (universities and laboratories), or funding/subsidising of private R&D activity. However, the growth effects of *public* R&D activities are empirically undetermined. If public and private R&D are seen as complements, then public R&D is expected to have a positive effect on growth[56]. If public and private R&D are seen as substitutes, then public R&D is expected to crowd out private R&D[57]. Currently

[54] Englander and Gurney (1994); De Gregorio (1996); Keefer and Knack (1997); De la Fuente and Domenech (2000); Bassanini and Scarpetta (2001); Gemmell and Kneller (2001); Heitger (2001); and Buysse (2002).
[55] Bleaney, Kneller and Gemmel (2001); Bloom, Canning and Sevilla (2001).
[56] Robson (1993); Park (1995); Busom (1999); Diamond (1999); Guellec and van Pottelsberghe de la Potterie (2000).
[57] Toivanen and Niininen (1998); Wallsten (1999); Bassanini, Scarpetta and Hemmings (2001).

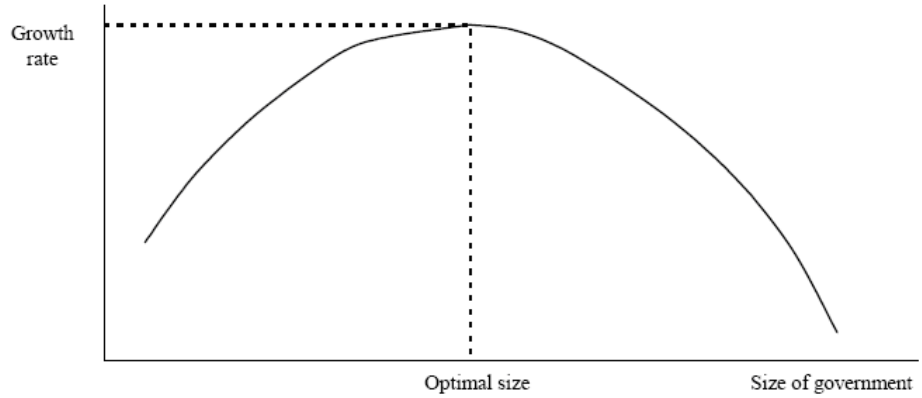
there is an equal balance of views on these two opposing perspectives, so the question remains unsolved[58].

Social expenditure on basic social safety nets is in theory expected to reduce the need for precautionary savings, improve the ability for risk taking and consequently act as another growth-promoting institutional factor (Afonso, Ebert *et al.*, 2005). However, pure redistributive spending beyond a certain scale can undermine growth by exerting a negative effect on incentives to work, and invest in human capital and in entrepreneurial enterprise.

Most econometric studies on the effects of general government spending on economic growth find a negative relationship between the two variables[59]. The theoretical arguments for such relationships are rooted in various soils: superfluous policy-induced distortions and the associated tax burden; crowding-out of private investment; and inefficient resource use due to insufficient market forces (Dar and AmirKhalkhali, 2002).

Some studies (Barro, 1990; Heitger, 2001), however, suggest the relationship between government size and economic performance is non-linear, and that a negative correlation between the two is likely to emerge only when government size exceeds a given optimal size (see Figure 3.1).

Figure 3.1: Economic growth and size of government (adapted from Gwartney *et al.*, 1998, exhibit 2)



[58] The relative merits of one view or the other seem to critically depend on the particular field of R&D activities considered as well as on the incentive mechanism provided (European Commission, 2004).
[59] See for example Barro and Sala-i-Martin (1995); Bassanini, Scarpetta and Hemmings (2001); De Gregorio (1996); Fölster and Henrekson (1999); Heitger (2001); Lee (1995).

Barro (1990) argues that an increase in public expenditure is likely to increase the marginal productivity of capital, while an increase in taxes is likely to reduce it[60]. The first positive effect is said to dominate when government is small, being offset by the tax effect when it grows beyond a certain point. Economic performance is then defined as an inverted U-shaped function of government size. Heitger (2001) agrees with Barro and attributes the downward part of the inverted-U function partly to the negative tax effects and partly to crowding-out effects. Gwartney, Lawson and Holcombe (1998) reach similar conclusions, arguing that government is likely to exert an overall positive effect on economic performance as long as it is confined to some core activities[61]. As government activity exceeds these core functions their net effect is expected to be adverse to economic growth because of disincentives caused by high taxation; diminishing returns as governments expand their activities into areas for which they are ill-suited; and interference with the market mechanism of allocation.

Studies portraying economic performance as an inverted-U function of government size suggest that in many of the countries analysed governments have outgrown both the scope of activities and the level of spending in the core activities mentioned above. Increasing the size of government in those countries is therefore expected to have a negative impact on growth as the crowding-out of private investment and other distortions caused by high taxation are likely to become dominant.

The legitimacy of these conclusions is mitigated by the fact that the size-effect analysis underlying these studies does not take into account the *efficiency* of public spending, which is an additional factor that mediates the relationship between government size and economic growth. Apart from being limited to a set of core activities, and within certain levels, public expenditure needs to be efficiently administered if it is to have a positive effect on economic growth.

3.7 Public Service Reform: from Hierarchies to Markets

[60] There is econometric evidence that a high level of taxation damages the general allocation of resources by distorting incentives to work, to invest and/or to save - Cashin (1995); de la Fuente (1997); Folster and Henrekson (1999); Kneller, Bleaney and Gemmell (1998). Growth is also impaired by high and sustained government deficits and growing debt (Tanzi and Chalk, 2000).

[61] Including the protection of persons and property; national security; education, physical infrastructure; and environmental protection.

In most developed nations the share of public spending and taxation in the economy rose steadily until the end of the 90s. Since then, a growing concern with the potentially negative consequences of such a trend has prompted a widespread adoption of various initiatives aimed at curbing the enlargement of the public sector, or at least bringing its growth to a halt (Joumard, 2002).

As we have just seen, this general movement to scale down the size of government is based on two fundamental theoretical/ideological assumptions: first, that high levels of public expenditure and taxation have distorting consequences on private incentives and therefore on economic performance; and second, that economic growth and productivity are hampered by high levels of public spending and taxation since the resources consumed by the public sector would have been more efficiently administered by the private sector.

In addition to these theoretical arguments, pressures to scale down government have also arisen from more pragmatic reasons. In the EU the monetary union and the Stability and Growth Pact have set restrictive norms on public debt and deficits. High unemployment rates have created pressures to reduce labour taxation. Population ageing has compelled governments to consolidate public finances in anticipation of rising financing burdens with health and pension expenditures. Finally, the global integration of markets and the inherent mobility of capital have eroded the capacity of individual nations to maintain high levels of taxation without compromising their international competitiveness.

The need to scale down the size of government has become the main driver of most public sector reforms, putting pressure on governments to re-conceptualise the form of their intervention in the economy and increase the efficiency of their activities. In the previous chapter we concluded that different modes of governance yield different levels of efficiency depending on the attributes of transactions and the scale of coordination required – i.e. the impact of different organisational arrangements (including different degrees of public sector involvement in the funding and provision of services/activity) on efficiency depends on the type of service/activity under consideration.

In the light of that result, we now hypothesise that improving the efficiency of government activity implies that (i) governments retreat from interfering with transactions that do not exhibit the kind of attributes that call for public central coordination; (ii) when public central coordination is required (e.g. education, health,

and social security), it should be confined to the specific attributes of transactions that markets fail to address[62] (public good attributes, merit good attributes, externalities, and/or information asymmetries), and in a way that explicitly acknowledges the risks of government failure[63]. An important result that follows from these two hypotheses is an expected positive effect of managerial decentralisation of service delivery units (public or private) on efficiency.

In practice, our research hypothesis involves re-defining the areas where government intervention is thought to be required (areas where the market left to itself is expected to produce sub-optimal results); re-defining the scope of such intervention (the extent to which public central coordination is needed, and in particular finding the right balance between the public and private sectors); and re-defining the particular forms whereby that intervention occurs (ranging from small regulation of private sector activity to the monopolistic provision of goods/services through centrally coordinated public bureaucracies).

Although the international experience of public sector reform is diversified (Pollitt and Bouckaert, 2004), there seem to be common elements that reflect these prescriptions to a certain extent. Some of these elements include the segregation of large public bureaucracies into smaller independent units with autonomy to operate relatively freely; the introduction of pricing and charging for public services; contracting-out services to the private sector on a competitive basis; control of public agencies exercised through performance-based regimes; development of property rights for managers as the basis of incentives; and the separation between the purchasers and providers of the services, i.e. the formal separation between the political and the operational dimension of collective action.

The public policy literature is crowded with arguments on the predicted effect these reforms *should* have on the efficiency of public services (e.g. Lane, 2000; Osborne and Gaebler, 1992; Lewin, 1991, Self, 1993; Udehn, 1996). However, none of these arguments is rooted in a comprehensive theory of economic organisation. Furthermore, research on the *actual* effect on efficiency is almost non-existent[64]

[62] Outside this domain, the coordination of transactions should still be decentralised as much as possible – by involving the private sector and/or by allowing state owned agencies enough flexibility to respond to market incentives.

[63] See section 4.3 of chapter 3.

[64] Some of the few existing evaluative empirical studies concentrate on the financial effects of contracting-out municipal services (solid-waste collection; bus transport; water supply; maintenance of vehicles, parks and streets, etc) to the private sector (e.g. Savas, 2000). However, these studies suffer

(Boyne, 1998; Pollitt *et al.*, 1998; Boyne *et al.*, 2003; Talbot, 2004). Most of the work on public service reform focuses on the causes of change (the economic, political and administrative antecedents of reform) rather than on its effects (Leemans, 1976; Caiden, 1988; Zifack, 1994; Pettigrew *et al.*, 1992; Lane, 1997; Kickert, 1997; Olsen and Peters, 1998; Pollitt and Bouckaert, 2000).

In the second part of the thesis we draw on the experience of OECD countries in the provision of education, health and social protection services to test our hypothesis of the effect of modes of governance on efficiency. Education, health, and social protection represent the core functions of the welfare state. These are services where reforms advanced the most and where the most audacious initiatives at reformulating the role of the public sector took place. The three services alone account for an average of roughly 60% of public expenditure in OECD countries (28% of GDP). The level of efficiency in the provision of these services is thus crucial to the overall efficiency of the public sector and to economic performance.

There has been recent progress in measuring efficiency in some of these areas (e.g. Afonso and St. Aubyn, 2004, for health and education; Afonso, Tanzi and Schuknecht, 2005, for the public sector as a whole). However, as far as we know, no attempts have been made so far to investigate the effects of different modes of governance on the efficiency of these services. In the absence of a proper analytical framework for measuring efficiency and evaluating the impact that shifting modes of governing public services have had on efficiency, the discussion around the effects of public sector reforms on efficiency is informed not by scientific knowledge, but crude ideological taste (Dunleavy and Hood, 1994[65]).

3.8 Conclusion

NIE compares the competence of different modes of organising transactions to take advantage of the division of labour while simultaneously economising on *specific* transaction costs caused by opportunistic behaviour and noncontractible uncertainties.

---

from important limitations both in terms of the economic significance of the services scrutinised, the context-specific nature of the data used, and the simplistic financial (cost-oriented) perspective adopted.

[65] "Most supposedly empirical discussions of the complex issues involved are dominated either by new public management evangelists exaggerating the efficiency impacts of changes on the basis of very preliminary or selective data; or by detractors basing their scepticism on dramatic anecdotes or sketchy arguments from past experience" (Dunleavy and Hood, 1994, p.13).

Based on a narrow conception of transaction costs, and a sole focus on production[66], NIE fails to provide an in-depth account of the inner-workings of alternative modes of governance that explain differences in aptitude to organise transactions efficiently.

Resource-based/evolutionary perspectives of the firm bring to light the limitations of NIE by drawing attention to the importance of interrelationships between activities (and associations between the capabilities that support them) in explaining the virtues of hierarchical coordination. However, these theories also fail to provide a comprehensive account of the link between organisation and efficiency.

This chapter provided an analytical framework for the study of economic organisation that brought together NIE and resource-based theories of organisation under a broader conception of transaction costs - all the costs of running a governance mechanism, including costs of coordinating and motivating individual/organisational decisions (Milgrom and Roberts, 1992).

Following the work of Friederich Hayek, we have conceptualised economic organisation as a means of bringing order into systems of human interaction. Order can be accomplished through a multitude of modes of governance - a continuum ranging from hierarchies to markets, i.e. from conscious deliberate coordination to spontaneous self-reinforcing coordination of self-interested individual actions. Transaction costs of the kind defined by NIE and associations between activities and capabilities of the kind defined by resource-based theories of organisation are an important part of this continuum, but they do not define it *per se*.

More broadly, we have demonstrated that decentralised coordination of transactions with *design* and/or *innovation attributes* (Milgrom and Roberts, 1992) is sub-optimal due to the brittleness of the system in dealing with imperfect information and excessive demands in terms of communication. For such types of transactions, hierarchical coordination is optimal.

In parallel, decentralised coordination of transactions with public good attributes, externalities, merit good attributes, and/or information asymmetries is also shown to be sub-optimal and to lead perfectly competitive markets to produce inefficient outcomes. *Public* central coordination is required to address these features and generate an efficient outcome.

[66] Transactions between consumers and producers are kept out of the analysis. The focus is exclusively on productive issues, and in particular on the make-or-buy decision (producer-supplier relationship versus in-house production).

The overlap of these two principles formed the basis of our understanding of the link between governance and public service efficiency reflected in our main research hypothesis: optimising government efficiency requires that (i) governments retreat from interfering with transactions that do not exhibit the kind of attributes that call for public central coordination; (ii) when public central coordination is required (e.g. education, health, and social security), it should be confined to the specific attributes of transactions that markets fail to address (public good attributes, merit good attributes, externalities, and/or information asymmetries), and in a way that explicitly acknowledges the risks of government failure.

The second part of this research hypothesis implies that the coordination of attributes that do not require public central coordination should be decentralised as much as possible, either by allowing the private sector to operate freely in a competitive environment and/or allowing state owned agencies enough flexibility to respond to market incentives.

In summary, we argued in this chapter that (in)efficiency in the provision of public services depends on the way key organisational design variables (in terms of funding and provision arrangements) interact between themselves and with the specific attributes of each service to create incentives for efficient behaviour. This principle is systematically deployed in chapters 5 through 7 to formulate testable research hypotheses on the effects of governance on efficiency in the context of primary and secondary education, healthcare, and social protection, respectively. Before that, the next chapter focuses on the metrics of public service efficiency.

Chapter 4: Research Design

4.1 Introduction

This chapter concentrates on the technicalities of efficiency measurement and the peculiarities and difficulties of measuring public service efficiency.

The methodological issues involved in the measurement of efficiency are presented in three stages - when inputs and outputs are defined and measured; when the set of feasible input-output combinations is estimated (i.e. the production/efficiency function/frontier); and when *actual* input-output combinations are compared with *feasible* input-output combinations.

Four sub-sections are devoted to presenting the most important methodologies associated with the estimation of the set of *feasible* input-output combinations (Data Envelopment Analysis; Free Disposable Hull; and Stochastic Frontier Analysis) and discussing their relative merits.

The final sections of the chapter concentrate on the specific methodological issues concerning the measurement of *public* service efficiency.

4.2 Efficiency Measurement: methodological issues

As proposed by Farrell (1957)[67], *economic* efficiency comprises two different components: *technical* efficiency, which reflects the ability of a decision-making unit (DMU)[68] to obtain maximal output from a given set of inputs; and *allocative* efficiency, which reflects the ability of the DMU to use inputs in optimal proportions and produce the right mix of outputs (given the prevailing prices and the production technology). Technical efficiency (TE), allocative efficiency (AE), and economic efficiency (EE) are put together in the overall efficiency relation[69]: $EE = TE * AE$.

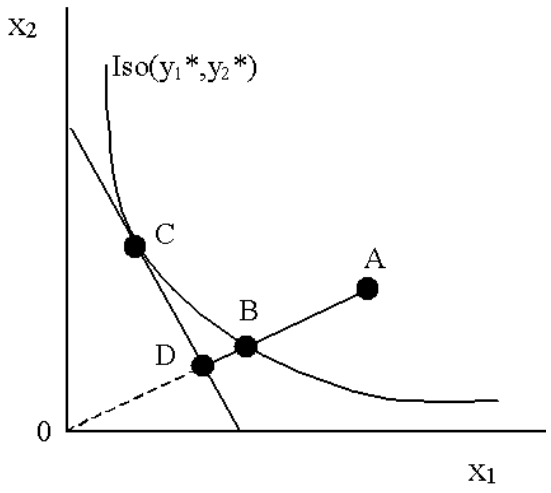
Efficiency is open to input/output interpretations depending on the specific orientation adopted (optimal combination of inputs to achieve a given level of output – input orientation; or optimal combination of outputs that could be produced with a

[67] Following the work of Debreu (1951) and Koopmans (1951).
[68] Any entity that directly or indirectly transforms inputs into outputs (for-profit firms, non-profit organisations, public organisations, etc).
[69] Measurement of technical efficiency is sometimes preferred to overall economic efficiency. Technical efficiency is a purely physical notion that can be measured without recourse to price information (often difficult to find and/or flawed), and without having to impose a behavioural objective on producers (cost minimisation; revenue, or profit maximisation).

given set of inputs – output orientation. Input/output-oriented measures of efficiency are only equivalent when constant returns to scale exist (Fare and Lovell, 1978).

Figure 4.1 illustrates theses concepts in a simplified two input (x_1, x_2) - two output (y_1, y_2) production process.

Figure 4.1: Input efficiency

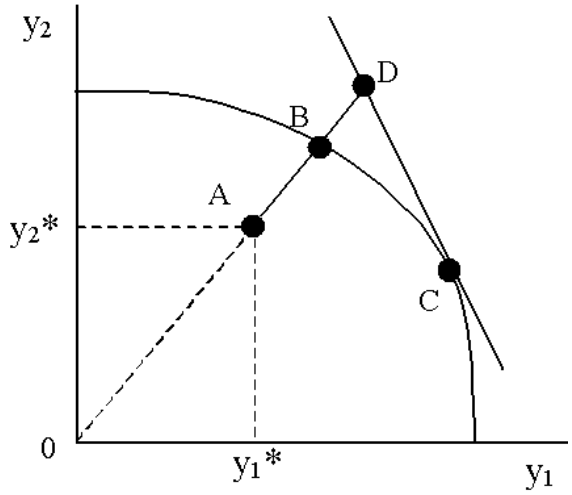


Suppose that the DMU produces a given level of output (y_1^*, y_2^*) using the input combination defined by point A. The same level of output could have been produced with a radial contraction of both inputs (point B), which lies on the isoquant defined by the minimum level of inputs required to produce (y_1^*, y_2^*). Therefore, the input-oriented level of technical efficiency ($TE_I(y,x)$) of the DMU is given by the ratio 0B/0A.

The minimum-cost combination of inputs that produces (y_1^*, y_2^*) is given by point C (where the marginal rate of technical substitution between the two inputs is equal to the input price ratio w_2/w_1). To achieve that level of cost, both inputs would need to be further contracted to point D. Input allocative efficiency ($AE_I(y,w_1,w_2)$) is therefore given by 0D/0B (Kumbhaker and Lovell, 2000). Overall economic efficiency $EE_I=(TE_I(y,x)).AE_I(y,w_1,w_2)$= (0B/0A).(0D/0B)=OD/OA.

Figure 4.2 illustrates the production possibility frontier for a given set of inputs and allows us to measure efficiency in an output-orientated way.

Figure 4.2: Output efficiency



Once again, it is assumed the DMU currently operates at point A. Technically efficient production using the inputs currently employed by the DMU would radially expand output to point B. Hence, the output oriented measure of technical efficiency ($TE_O(y,x)$) = 0A/0B.

Although point B is technically efficient (it lies on the production possibility frontier), the DMU could generate higher revenue (price-weighted output aggregate) with the same level of inputs by producing at point C (where the marginal rate of transformation between y_2 and y_1 equals the price ratio p_2/p_1). To achieve the revenue associated with point C while maintaining the same output combination, the DMU would have to expand output to point D. Hence, output allocative efficiency ($AE_O(y,p_1,p_2)$)=0B/0D (Kumbhaker and Lovell, 2000). Finally, overall economic efficiency EE_O=($TE_O(y,x)$)* ($AE_O(y,p_1,p_2)$)= (0A/0B)* (0B/0D)=OA/OD.

In practice measuring efficiency usually involves three sequential steps. First, inputs and outputs need to be properly defined and measured. Secondly, the set of feasible input-output combinations needs to be estimated (i.e. the production/efficiency function/frontier[70]). Finally, comparing *actual* input-output combinations with *feasible* input-outputs yields efficiency estimates.

[70] The technical relationship between inputs and outputs is usually designated *production/efficiency frontier* in the efficiency measurement literature, and *production function* in orthodox microeconomic textbooks.

Although the first two steps are relatively straightforward in a one-input/one-output scenario, substantial analytical difficulties often arise when the technology involves multiple inputs/outputs, particularly in the context of public services. An example of such a difficulty, which we will discuss later in greater detail, arises in the valuation of outputs. While in the case of private organisations operating in competitive markets different outputs can be aggregated using observed prices as weights, these are all but absent in most public services, which often makes it very difficult to estimate the value of the output of these services. Another difficulty concerns the treatment of variables that are capable of affecting efficiency but that lie outside organisations' managerial control. The way in which the effect of these so-called environmental variables should be allowed for in measuring efficiency is often controversial (Coelli *et al.*, 1998).

There are two distinct approaches to estimating the efficiency frontier. The first involves the theoretical definition of the set of feasible input allocations, their costs and associated outcomes (under conditions of technical efficiency), as well as the identification of the set of outputs that maximise the objective function for the resources available. This approach is extremely complex and demanding in terms of the information required, rendering it unfeasible for international comparative studies.

The second approach is to estimate the frontier from a sample of observed inputs and outputs from different DMUs. The procedure is to compare a DMU with an ideal comparator constructed from information on other DMUs operating in the same field (and ideally with similar size and environmental factors). This provides a relative benchmark or comparison among DMUs, on the basis of the real or observed performance of units, and not the theoretical maximum derived from an abstract theoretical production function as in the first approach.

Within the second approach, Farrel (1957) suggested that the efficiency frontier be estimated either by (a) a non-parametric piece-wise-linear convex isoquant, or (b) a parametric function fitted to the data, both constructed such that no observed point lies to the left or below them. Since this seminal work, research on efficiency measurement has diverged into two streams, with economists typically following the route of statistical analysis and management scientists typically opting for a non-parametric route grounded in linear programming.

The *non-parametric route* originally explored by Charnes, Cooper and Rhodes (1978) resulted in the development of the Data Envelopment Analysis (DEA). DEA uses linear programming methods to construct a non-parametric piece-wise surface (or frontier) over the data, assuming free disposability and convexity of the production set. A more general nonparametric approach is Free Disposable Hull (FDH), introduced by Deprins, Simar and Tulkens (1984), which assumes only free disposability of the production set.

The *parametric route* initially introduced by Aigner and Chu (1968) is based on a deterministic[71] model using a Cobb Douglas functional form. This model did not allow for the influence of measurement errors or statistical noise on the frontier. All deviations from the frontier were assumed to be the result of technical inefficiency.

Another early application of parametric efficiency analysis in which the entire residual was interpreted as being due to inefficiency is the so-called corrected ordinary least squares (COLS) approach. A production function is first estimated using ordinary least squares (OLS). The OLS intercept parameter is then shifted up by the value of the largest positive residual to give the equation for the frontier. Again, no allowance is made for the possibility that the residual may incorporate factors other than inefficiency, such as measurement error or omitted model variables.

As a means of tackling some of the limitations of these methodologies Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977), proposed the stochastic frontier production function (SFA), in which the residual is separated into two components with zero covariance: inefficiency, u_i, and all other sources of error, v_i.

The following sections describe the most relevant aspects of DEA, FDH and SFA in greater detail.

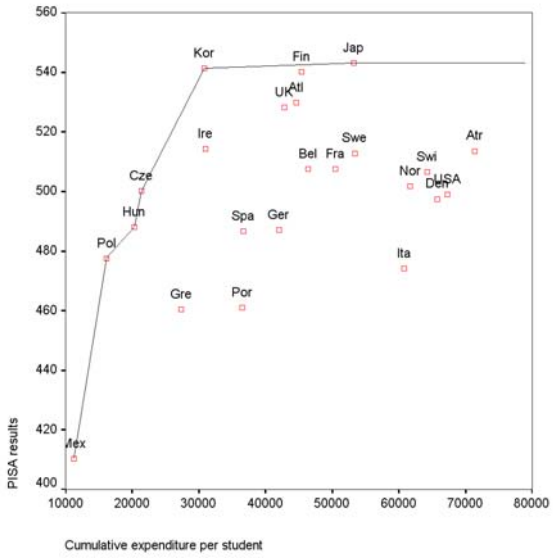
4.2.1 Data Envelopment Analysis

Data Envelopment Analysis originates from Farrell's (1957) seminal work and was popularised by Charnes, Cooper and Rhodes (1978). It assumes the existence of a convex production frontier and uses linear programming methods to construct linear

[71] Observed output is bounded above by a non-stochastic term.

segments that interpolate between those observations with the highest ratios of output to input (e.g. Figure 4.3).

Figure 4.3: DEA frontier



Assuming just one input X, the DEA formulation of the production process seeks for each organisation "0" to find a set of weights $\boldsymbol{w}^0$ and an efficiency score θ_0 so as to:

$$\max\ \theta_0 = \frac{\sum_{r=1}^{s} y_{r0} w_r^0}{X_0}$$

$$\text{subject to } \frac{\sum_{r=1}^{s} y_{r0} w_r^0}{X_j} \leq 1 \qquad \text{j} = 1,\ldots,\text{n}$$

Equation 4.1

where $\boldsymbol{w}^0$ is the set of weights that maximizes the performance of organisation 0, subject to requiring that, using that set of weights, no organisation has an efficiency score greater than 1. In other words, it searches for a linear composite of other organisations that produces the same (or more) of each output (y_r), but at the least cost a fraction $1/\theta_0$ of the costs X_0 of organisation 0 (Smith and Street, 2005).

A separate linear program is estimated for each organisation, and therefore a different set of weights $\boldsymbol{w}^j$ will be computed for each organisation. The efficiency score thus obtained for each observation uses the set of weights that is most favourable for that observation.

In a multiple input/output context, the fractional programme equivalent to (Equation 4.1) can be reformulated into a linear programme by constraining the numerator or denominator to be equal to unity. The optimal values of θ_0 and $\boldsymbol{w}^0$ can then be determined using standard linear programming techniques. The input-minimising programme is:

$$\begin{aligned} &\min_{\theta,\lambda} \theta, \\ &\qquad -y_i + Y\lambda \geq 0, \\ &\text{subject to} \quad \theta x_i - X\lambda \geq 0, \\ &\qquad \lambda \geq 0, \end{aligned}$$

Equation 4.2

where θ is a scalar (the efficiency score of DMU i) and λ is a vector of constants. For each DMU, the programme seeks to maximize a radial contraction of the vector of inputs (x) constrained to remain within the feasible input set.

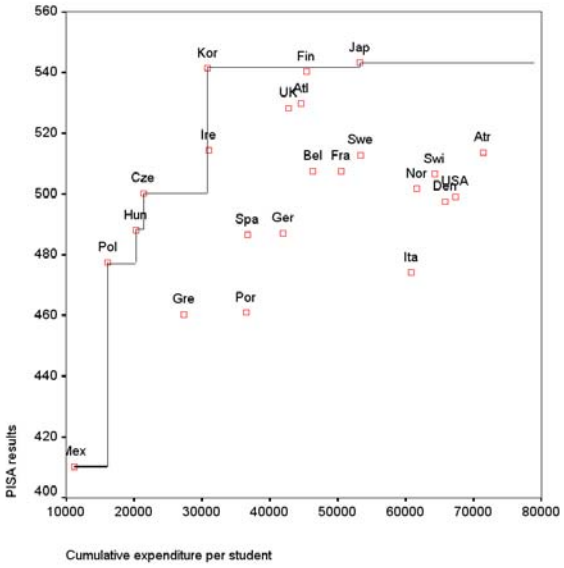
The linear programme (Equation 4.2) corresponds to the original Charnes, Cooper and Rhodes (1978) formulation of the DEA model assuming constant returns to scale (CRS). A variable returns to scale (VRS) frontier can be obtained by adding to programme (Equation 4.2) the constraint $N1'\lambda = 1$, where N1 is a vector of ones (Banker, Charnes and Cooper, 1984). This second formulation forms a convex hull of projecting planes enveloping the data more tightly than the CRS conical. This ensures each DMU is effectively evaluated against DMUs of similar size.

Recently, some of the most interesting extensions of this basic DEA model have included the use of panels of observations and bootstrapping efficiency estimates to assess statistical significance (Thanassoulis, 2001; Coelli *et al*., 1998; Simar, 1992; Simar and Wilson, 2000, 2002).

4.2.2 Free Disposable Hull

Free disposable Hull (FDH), first formulated by Deprins, Simar, and Tulkens (1984), has nearly identical modelling features and properties to DEA except that it omits any assumptions concerning convexity and/or returns to scale. The basic motivation is to ensure that efficiency evaluations are derived only from observed performances. The production possibility set is defined by all the points whose input coordinates are at least as large as their corresponds in the vector of observed values x_j for any $j=1,...,n$, and whose output coordinates are no greater than their corresponds in the vector of observed values y_j for any $j=1,...,n$ (Cooper, *et al.*, 2000). The resulting efficiency frontier is a step function of the type portrayed in Figure 4.4.

Figure 4.4: FDH frontier



The following mixed integer linear program defines the FDH production frontier:

$$\min_{\theta,\lambda} \theta,$$

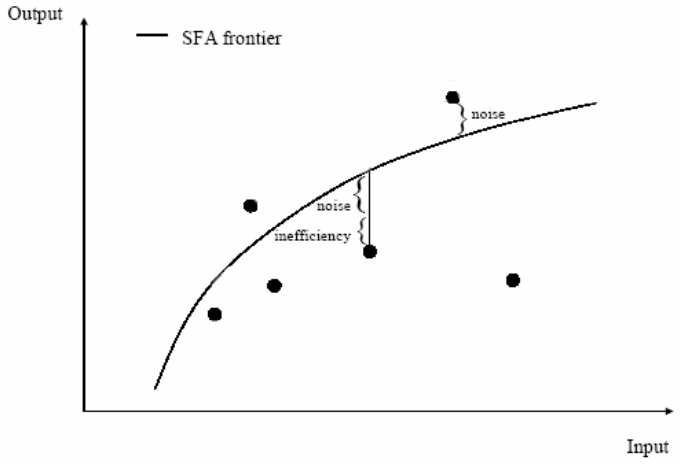
$$\text{subject to} \quad \begin{aligned} & -y_i + Y\lambda \geq 0, \\ & \theta x_i - X\lambda \geq 0, \\ & N1'\lambda = 1, \\ & \lambda_j \in \{0,1\}, \end{aligned}$$

Equation 4.3

4.2.3 Stochastic Frontier Analysis

The stochastic specification of the production function frontier was independently proposed by Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977). This specification explicitly acknowledges that some deviations from the maximum observed output may occur due to factors unrelated to inefficiency (e.g., inaccuracy in the measurement of output; exogenous shocks outside the control of the production system, etc; see Figure 4.5).

Figure 4.5: SFA frontier



A parametric production function is estimated assuming the error term has two components: one representing random errors v_{it}, and the other representing technical inefficiency u_{it}. Let Y_{it} denote output of country i in time period t, the production function is of the following form:

$$Y_{it} = \alpha + X'_{it}\beta + v_{it} - u_{it}$$

Equation 4.4

where X_{it} is a vector of inputs and v_{it} is the error term with mean zero. The term $u_i \geq 0$ is a random variable representing DMU-specific technical inefficiency, and is constrained to be always non-negative. Technical efficiency is then defined as the ratio of the expected value of observed output for country *i* to the expected value of the output when $u_i = 0$:

$$TE_i = \frac{E(Y_{it}|u_i, X_{it})}{E(Y_{it}|u_i = 0, X_{it})}$$

Equation 4.5

The denominator represents the frontier since the technical inefficiency term takes the value of zero. When Equation 4.4 is applied to cross-sectional data the distributional form of the two residual components v_i and u_i must be specified explicitly and maximum likelihood techniques used to estimate the model parameters. The random error v_i is usually modelled as the traditional two-sided error with a normal distribution. The inefficiency error component u_i is modelled as a one-sided distribution as inefficiency is reflected in a negative deviation from the efficient frontier (Schmidt and Lin, 1984). Various distributional assumptions for u_i have been used in empirical applications of Equation 4.4, including the half-normal, truncated normal, exponential and gamma distributions. There is no theoretical basis for favouring one distribution over another (Schmidt, 1985).

When Equation 4.4 is applied to longitudinal data, an organisation-specific fixed or random effect can be estimated and interpreted as a measure of inefficiency. Recently, some alternative estimation procedures have considered the possibility that fixed or random effects may incorporate unobserved organisation-specific influences on performance other than inefficiency (Greene, 2005; Farsi *et al.*, 2003).

4.2.4 Methodological Comparative Discussion

The relative merits of each of these approaches to the estimation of the efficiency frontier are well documented (see for e.g. Henderson, 2003).

Within the non-parametric paradigm, apart from differences in DEA models following the various possible assumptions regarding returns to scale, there are substantial differences between the DEA and FDH techniques that should be taken into account when comparing results.

The linear programming technique used in DEA measures efficiency as the distance from a convex envelope of the data. An undominated input/output set is deemed inefficient if it does not lie on the convex envelope. In contrast, FDH is essentially concerned with dominance rather than distance. Only input/output combinations dominated by other observed input/output combinations are deemed inefficient. Some authors (e.g. De Borger and Kerstens, 1996; and Vanden Eeckaut, Tulkens and Jamar, 1993) argue the convexity assumption fails to recognize local nonconvexities, leading to an overestimation of inefficiency. DEA models often compare inefficient observation to unobservable, fictitious linear combinations of efficient observations. FDH does not suffer from this problem as it relates each inefficient observation to a single dominating observation. However, in small samples it is likely to underestimate inefficiency due to the lack of comparable DMUs[72].

The relative merits of the parametric and non-parametric paradigms are also well established (Coelli *et al.*, 1998):

- Non-parametric approaches assume all deviations from the frontier occur exclusively on account of inefficiency, which renders them highly vulnerable to measurement error, exogenous shocks and omitted variables; furthermore, the non-stochastic character of these approaches does not permit testing hypotheses regarding the inefficiency and the structure of the production technology.

- Stochastic frontiers also have some significant limitations of their own. The selection of the distributional form for the inefficiency effects is often arbitrary, and it may have a significant impact on the efficiency estimates obtained; the production technology needs to be specified *a priori* through a particular functional form;

[72] As Henderson (2003, p. 7) notes, "in the FDH framework, an observation with epsilon amount less of a particular input and a substantial amount less of output than an efficient firm may be deemed efficient, whereas that firm would be considered to be highly inefficient by DEA".

multiple-output technologies imply the aggregation of output into a single measure, or the assumption of cost-minimising behaviour.

Gong and Sickles (1992) use Monte Carlo experiments to compare DEA with several econometric techniques. The performance of the estimators is assessed using the correlation and rank correlation coefficient between the true and estimated inefficiencies, which measure the degree to which deterministic and stochastic efficiency estimates are in accordance with the true data generating process. The authors show that the stochastic model outperforms the DEA model, when the functional form employed is close to the underlying technology. However, the choice of functional form proves to be important in deriving unbiased information about the firm-specific technical efficiency. Furthermore, the experiments suggest the tracking ability of both the deterministic and stochastic techniques deteriorates rapidly as the true technology becomes more complex.

In chapters 5 through to 7, the stochastic specification of the production function is our generally preferred method of estimating efficiency[73]. Four factors motivate our choice: (i) our analysis is generally confined to single-output technologies, which immediately rules out favouring non-parametric approaches on the basis of their *alleged* relative advantage in dealing with multiple outputs; (ii) the stochastic approach in its own nature to test hypotheses regarding inefficiency and the structure of technology, whereas the adjustments to non-parametric approaches that allow such hypotheses testing are seldom exempt of important limitations; (iii) the stochastic approach is less prone to be influenced by measurement errors, exogenous shocks, and omitted variables; (iv) the need to specify *a priori* the production technology (often presented as one of the main limitations of the stochastic approach) can be mitigated by the adoption of flexible functional forms (e.g. translogarithmic production function).

4.3 Efficiency vs Productivity

Most of the literature on the effects of public service reforms has concentrated on changes/differences in productivity (e.g. Pollitt *et al.*, 1998; Boyne *et al.*, 2003;

[73] The exception is chapter 6 where data envelopment analysis is used in parallel with stochastic frontier analysis as a means of testing the robustness of results in the context of multiple-output technology.

Talbot, 2004). However, there are problems in interpreting productivity growth/differences as welfare or efficiency measures.

Differences in productivity between different DMUs or individual DMUs in different moments in time, depend on dissimilarities/changes in efficiency, technical change, exploitation of scale economies, or some combination of these three factors (Coelli *et al.*, 1998).

Let y_t, y_s, and x_t, x_s represent observed quantities of outputs and inputs produced by a DMU in periods t and s, respectively. Let the production technologies in these periods be represented by $f_s(x)$ and $f_t(x)$. The productivity change between periods s and t is then defined by the total factor productivity (TFP) index[74]:

$$TFP_{st} = \frac{y_t / y_s}{x_t / x_s}$$

Equation 4.6

If the DMU is technically inefficient in both periods t and s, and assuming output efficiency is reflected in the scalar λ then[75],

$$y_j = \lambda_j f(x_j), \quad \text{where } 0 \le \lambda_j \le 1, \text{ j = s,t}$$

Equation 4.7

If we further assume that the level of input consumed is different between the two periods so that $x_t = \kappa x_s$, and that the production function is homogeneous of degree ε(t), at x_t in period t then[76],

$$TFP_{st} = \frac{\lambda_t}{\lambda_s} * \frac{f_t(\kappa x_s)/\kappa x_s}{f_s(x_s)/x_s} = \frac{\lambda_t}{\lambda_s} * \kappa^{\varepsilon(t)-1} * \frac{f_t(x_s)}{f_s(x_s)}$$

Equation 4.8

[74] Coelli *et al.*, 1998.
[75] *Ibidem.*
[76] *Ibidem.*

The first part of Equation 4.8 represents change in technical efficiency[77] (changes in the distance from the production function); the middle component reflects the effect of a change in the scale of operations (which depends on the scale of operations (κ) and the returns to scale parameter ε(t)); and the last component measures technical change (shifts of the production function)[78].

Cross-sectional and longitudinal comparisons of aggregate measures of productivity do not generally allow disentangling these three components unless very restrictive (non-credible) assumptions are made. Without a link between variations in productivity and changes or differences in efficiency, scale and technology, the analytical value of productivity variations is very limited. Dawson *et al.* (2004) provide a good illustration of this point (Figure 4.6).

Figure 4.6: Productivity and efficiency (adapted from Dawson *et al.*, 2004, p. 7)

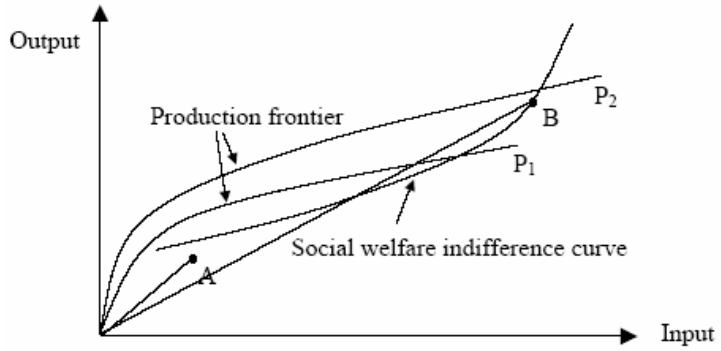


Figure 4.6 depicts a simple single input, single output technology. Point A in year 1 has higher productivity than point B in year 2 (given by the slope of the segments that link each point to the origin). However, both welfare and technical efficiency are lower at point A (A is below the social welfare indifference curve and is further from its period production frontier). Technical progress shifted the frontier

[77] In a more general formulation, the valuation of inputs and outputs would introduce "allocative efficiency" as an additional component to equation 4.8.

[78] Note that if as an alternative to the derivation of TPF index 4.8 we use period-s level of input $x_s=\mu x_t$ then $TFP_{st} = \frac{\lambda_t}{\lambda_s} * \mu^{1-\varepsilon(s)} * \frac{f_t(x_t)}{f_s(x_t)}$. The measure of technical change given by the last component in this equation is only equivalent to that of equation 4.8 if there are constant returns to scale (in the single-input case) and input homotheticity in the case of multi-input/multi-output production. Also, if the TPF index was derived using an input oriented approach instead of the output oriented approach implicit in equation 4.8, the technical change and technical efficiency measures of the two approaches would only coincide in the presence of CRS/input homotheticity (Coelli *et al.*, 1998).

upwards from P_1 to P_2, but productivity change does not even have the same sign as technical progress. The increase in welfare between period 1 and 2 is partially explained by technical progress (B was not feasible under the old technology) and partially explained by improvements in efficiency. Although both technologies in this example have diminishing returns to scale (increases in inputs along the frontier reduce productivity), moving along the frontier can be welfare increasing.

In other words, declining productivity alone may not indicate reduced efficiency or welfare but simply rapid increases in expenditure coupled with diminishing returns to scale[79]. These considerations suggest that there are problems in interpreting productivity growth as a welfare or efficiency measure. The growth in TFP is a combination of technical change, efficiency change (technical and allocative), and scale change.

4.4 Public Service Efficiency: Methodological Issues

As it has been mentioned in section 3, measuring efficiency involves three sequential steps: defining and measuring inputs and outputs; estimating the set of feasible input-output combinations; and comparing actual input-output combinations with feasible input-outputs combinations.

These steps are relatively straightforward in the case of private organisations operating in competitive markets. Even when organisations trade multiple inputs and outputs, prices are usually available for aggregating these operations and so efficiency can be easily estimated provided the number of observations in the sample is sufficiently large.

The provision of public services[80] though, is impregnated with peculiarities that hinder the process of measuring efficiency. Public services are normally provided free of charge (or at subsidised prices) at the point of delivery, which brings significant obstacles to the determination of their societal value and consequently to

[79] The expansion of public services may indeed be subject to diminishing returns. Historically, initial reductions in neo-natal mortality may have been relatively easy to achieve, but driving down rates today below their current level may involve much more expensive interventions. When schools have halved their truancy rate, they may find it increasingly difficult to achieve further reductions. When re-offending has been reduced by 5 per cent, the next 5 per cent may mean working with more difficult cases. Diminishing returns do not, of course, mean that the expansion is unjustified. The value of the additional output may still exceed the cost of the inputs.

[80] Irrespective of whether these services are funded and provided by government, or whether the private sector plays a major role in their provision.

the aggregation of their output. Public service inputs normally pose fewer problems, insofar as they are purchased in competitive markets, and prices are available for determining their relative value (just as in the case of the private sector). However, there are situations where government is not just another buyer in a competitive market. Instead, it may have considerable monopsony power; it may bargain with powerful trade unions; or prices may be determined by complex regulatory mechanisms. Sections 5.1 and 5.2 address the specificities of government outputs and inputs in the context of efficiency measurement.

Apart from difficulties in measuring inputs and outputs, there are a number of methodological choices relating to the estimation of the production frontier that are likely to have a significant impact on public service efficiency measurements. Some of the most important choices concern the determination of output weights; modelling the production process; controlling for environmental constraints; and allowing for dynamic effects (Smith and Street, 2005). Section 5.3 concentrates on these choices individually.

4.4.1 Output

Traditionally public sector output has been measured as of value equal to the total value of inputs used; by extension output volume has been measured by the volume of inputs. The wide spread usage of the "output=input" convention reflects difficulties of alternative estimates: the exact nature of the output generated by collective services (e.g. defence and public administration) is very difficult to define; and in the case of services supplied to individuals (e.g. education, health) the societal value of these services is also difficult to determine since there is no market transaction (Atkinson, 2005).

A useful distinction can be made between *activities*, *outputs* (goods/services which may require a bundle of activities), and *outcomes* (the characteristics of output which affect utility). Consumers value goods because of the bundle of utility yielding characteristics they produce. The *quality* of the output can also be seen as some function of the vector of outcomes it produces (Deaton and Muellbauer, 1980; Lancaster, 1971; Dawson *et al.*, 2004).

In competitive markets, prices measure the consumers' marginal valuations of the bundle of characteristics associated with the consumption of outputs. In the public sector, there are no final markets or prices to reveal consumers' marginal valuations of outputs, and so their value needs to be estimated. There are two ways of doing that: either measure the outputs and attempt to estimate the marginal valuations attached to them or measure the outcomes produced by each unit of output and attempt to estimate the marginal valuations of the outcomes (Dawson *et al*., 2004)

Let us assume the production technology is of the type $g_j(y_j, q_{j1}, ..., q_{jM}, z_j, v_j) = 0$, where y_j is the volume or quantity of output j (the number of units produced) and q_{jm} is the amount of outcome or characteristic m produced by consumption of one unit of output j. The vector $\mathbf{q}_j$ determines the quality of the product. At the equilibrium of a market economy the price paid for a unit of output j depends on the outcomes it produces $p_j(\mathbf{q}_j)$, and is also a measure of quality. The marginal effect of outcome m on the price of output j is $\pi_{jm} \equiv \partial p_j / \partial q_{jm}$. In the competitive equilibrium these prices represent social values as well as costs of production. Then if $p_j = \sum_m \pi_m q_{jm}$ we can replace the problem of estimating p_j with estimating the m marginal values π_m of the outcomes[81].

If the marginal social values of output (p_j) or of outcomes (π_m) vary over time or across DMUs, variations in the value of government output depend both on differences in production conditions (the rates of growth of outcomes per unit of output and the rates of growth of outputs) but also on differences in preferences (the rates of growth of the marginal social values of outcomes). Changes in the value of outcomes should affect decisions about the allocation of resources within the public sector and the relative size of the public sector, and therefore should be taken into consideration when efficiency is being measured.

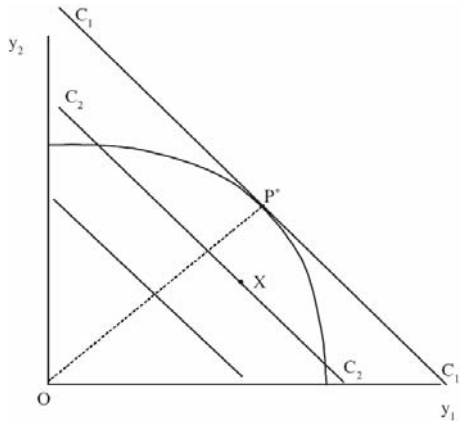
[81] Note that assumption that the marginal social value of a unit of output j is a linear function of its characteristics and that the π_{jm} is independent of j characteristics is strong. Dawson *et al.* (2004) provide an example from the health sector where this assumptions requires that an improvement in the quality of hospital food has the same effect on the value of treatment for throat cancer as on the value of a hip replacement.

In the past measures have been employed whereby the cost of public sector output is used to estimate its marginal value. The main problem with using costs to evaluate public output is the lack of a market mechanism ensuring that the marginal cost of public output equals the marginal value to consumers. It can be shown that cost weighted output indexes are equivalent to the value weighted index $p_j = \sum_m \pi_m q_{jm}$ only if (a) quality variation/change is zero for all characteristics of all outputs; and (b) costs are proportional to the marginal social value of output (Dawson *et al.*, 2004, section 2.11)[82].

An alternative approach to cost weights is to place judgements on the relative merits of different outputs by constructing a composite indicator. However, the use of "judgemental" weights is likely to be controversial at best and open to abuse at worst (O'Mahony and Stevens, 2004), and evidence suggests the composite is highly sensitive to these methodological choices (Jacobs, Smith and Goddard, 2003).

Composite indicators often take the form of linear performance indexes, which implicitly assumes that the trade-off between different outputs should be resolved by maximizing a linear function of the various output measures. Figure 4.7 provides an example with a simplified two-output technology. The parallel lines indicate different values of a chosen composite indicator, with scores increasing towards the top right-hand corner. Given the weights used in the composite indicator the choice P* on the possibility frontier is optimal, and the choice of any other point on the frontier would be allocatively inefficient (Smith and Street, 2005, p. 406, 408).

Figure 4.7: Production possibility frontier



[82] An additional limitation of using costs to evaluate public output is the difficulty in incorporating quality adjustments with cost saving properties.

4.4.2 Inputs

Measuring the inputs used by the public sector poses fewer conceptual problems as most inputs are sold on markets and prices are readily available. In competitive markets (where buyers and sellers are price takers and suppliers of inputs bear the full costs of their decisions) input prices reflect marginal social costs and thus can be used directly to value inputs. However, there are situations where government has considerable monopsony power; others where it bargains with powerful trade unions; and yet others where prices are determined by a complex regulatory mechanism (Dawson *et al.*, 2004).

In input markets where government purchases affect the market price, because the supply curve is upward sloping the marginal expenditure on the input will exceed the market price, and the amount bought will be less than if the government were price taker. Since suppliers will still be on their supply curve, and assuming the supply curve reflects all the costs of supply, the price in this monopsonised market will still equal the marginal social cost of the input.

In the opposite scenario where sellers have monopoly power, the price paid will exceed the marginal social cost of supply. This is also likely to be true in markets where government bargains with powerful trade unions. The precise levels of wage and employment will depend on the specification of the bargaining model (e.g. whether it occurs over the wage and employment, or just over the wage, with government acting as price taker in choosing employment levels). In either case, however, the bargained wage will include a rent and so will exceed the marginal social cost of labour.

When prices are determined by regulatory mechanisms (e.g. the 1999 Pharmaceutical Price Regulation Scheme in the health sector), the suppliers' rate of return on inputs sold to the public sector is generally limited. The effect of regulation is thus to push prices closer (on average) to marginal production costs, although there will tend to remain a gap between the two. In the pharmaceutical market for example prices are sufficiently above marginal production costs to enable firms to earn a rent which is expected to provide proper incentives for research and development of new drugs.

4.4.3 Efficiency Measurement

In addition to problems in defining and measuring public service outputs and inputs, public service efficiency measurements are sensitive to the specific set of methodological choices adopted in the estimation of the productive frontier.

Smith and Street (2005) lay emphasis on the consequences of four of these choices: the choice between a parametric or non-parametric approach to the estimation of the production frontier imposes distinct assumptions on output weights; decisions on the specification of the production process are likely to have an important impact on the efficiency estimates; the set of environmental variables to be included in the model, as well as the way in which these variables are controlled for is contentious irrespective of the particular approach adopted for the estimation of the frontier; finally, although allowing for dynamic effects in modelling the production process drastically increases complexity, this should be done whenever possible and the consequences of its disregard should be made explicitly from the outset.

4.4.3.1 Weighting

Measuring public service efficiency implies deciding on the set of *objectives* against which the performance of organisations will be evaluated and on the *weights* attached to these objectives. As we have seen in section 5.1, the set of weights **w** ought to reflect societal values. However, in practice determining vector **w** is often problematic. First, prices for valuing outputs or outcomes are generally non-available[83]. Second, public sector organisations usually face multiple objectives and lack a consensus on the prioritisation of those objectives. Ultimately, the selection of objectives and the determination of their weights should be the responsibility of politicians who are charged with reconciling conflicting claims on public resources (Smith and Street, 2005).

The parametric and non-parametric methodologies discussed in section 3 generate weights as a by-product of the estimation procedure, rather than

[83] The are various sources that could be used in an attempt to determine these prices including economic studies of willingness to pay, and use of prices of parallel markets, just to name a few.

incorporating a pre-defined vector of weights[84]. Some authors actually see this as an attractive feature of these methods (Cooper *et al.*, 2000).

In the parametric paradigm, estimating efficiency with multiple outputs is usually accomplished with recourse to either cost functions[85] or distance functions (Coelli and Perelman, 2000; Lothgren, 2000; Shephard, 1970). In both cases, the estimated weight for each output (shadow price) corresponds to the value implicit in the sample mean cost of producing an additional unit of that output. The underlying assumption is that the observed expenditure choices of organisations (on average) reflect the values that are placed by society on the outputs.

In the non-parametric paradigm and particularly in DEA, the weights are allowed to vary freely so that each organisation's efficiency score is maximised. This is equivalent to allowing each unit to choose the criteria by which it wishes to be judged. This flexibility does not come without a cost though, as it undermines the conclusions that can be drawn about relative efficiency (Pedraja-Chaparro, Salinas-Jiménez and Smith, 1997). In effect, in small samples this may lead to some units being deemed efficient simply because they are different (in their input or output mixes) from other units[86]. A possible way of overcoming this problem would be to impose weights, but would bring us back to the problem of determining such weights removing the flexibility advantage of DEA.

4.4.3.2 Modelling the production process

Modelling the process whereby inputs are transformed into outputs plays a key role in efficiency measurement. Traditional empirical research models try to unveil the structure and determinants of the production process engrained in aggregate (sample average) patterns – e.g. the marginal productivity of factors of production; how the technology is characterised in terms of returns to scale; and how external environmental factors influence efficiency.

---

[84] As Smith and Street (2005, p. 409) note: "It is of course possible that the weights emerging from statistical studies correspond to political preferences. However, we are not aware of any studies that have sought to verify this. At the very least, we would suggest that there is a need for careful dialogue between policy makers and analysts to ensure that the methods that are used reflect policy requirements".

[85] This implicitly imposes a cost-minimising behavioural assumption on organisations.

[86] Related to this is the general issue of sensitivity in DEA. The relative efficiency score achieved by each unit can be sensitive to the number of inputs and outputs specified (Sexton, 1986; Nunamaker, 1985). The more input and output variables are included in the model, the higher will be the number of units with an efficiency score equal to unity (Nunamaker, 1985).

However, the managerial and policy interest often goes further in seeking to obtain information on the efficiency of individual organisations. Frontier analysis derives this information from the residual or organisation-specific effect. Focusing on the residual means the model parameters are no longer the main point of interest, which effectively "turns the statistical model on its head" (Smith and Street, 2005, p. 410).

Traditional statistical models seek to satisfy specific acceptability criteria[87]. Meeting these criteria ensures the models convey an adequate portrait of aggregate/structural features of production, but it does not guarantee they are appropriate for inferring efficiencies of individual organisations. In particular, individual efficiency can be under/overestimated as a result of model misspecification (omitted variables or functional form) or measurement errors. Some of these problems can be avoided by testing the underlying representation of the production technology through sensitivity analysis. In DEA, different scaling assumptions can be tested and bootstrapping estimates may be employed to assess statistical significance (Simar and Wilson, 2002). Stochastic frontiers allow for different functional forms and distributions of the inefficiency effects to be tested, and provide confidence intervals around estimates of inefficiency and of the model parameters.

In general, the sensitivity of results to different methodological choices depends on the complexity of the underlying production process. When this is process is relatively simple, the results are not expected to show wide variation across methods and assumptions. Smith and Street (2005) indicate estimates of the efficiency of companies providing water and sewerage services in the UK as an example. These appear to be relatively robust in the face of sensitivity analysis (Office of Water Trading, 1999). As a contrasting example, the authors mention the efficiency estimates derived from different models for the UK hospital sector, which seldom yield definitive or consistent conclusions (Jacobs, 2001).

4.4.3.3 Environmental constraints

[87] Typically these include consistency (as the sample size increases, estimates converge to parameters), unbiasedness (expected error in the estimate equal to zero), efficiency (sampling variance of the estimate as small as possible), robustness (sensitivity of estimates to potential model misspecification, missing information and measurement error) and parsimony (the model is as simple as possible).

The measurement of public service efficiency is further complicated by the influence on efficiency of variables that lie outside managerial control (so-called environmental variables) - e.g. characteristics of individuals being served; external environment (geography, climate, and culture); and activities of other related organisations.

Differences in these variables between organisations lead to differences in their production possibility frontiers (for each level of expenditure). The frontiers of organisations operating in less-favourable "environments" will lie inside those of more favourably endowed organisations. The effect of these variables needs to be taken into account when modelling efficiency otherwise efficiency will be over/underestimated.

The way in which these variables should be included in the efficiency models is not consensual (Ozcan, Luke and Haksever, 1992, Buck, 2000, Fried *et al.*, 2002). In DEA, they can be included alongside inputs in the production model, ensuring each organisation is compared only with organisations operating in identical or less favourable environments. Alternatively the model can be first estimated ignoring the environment, and in a second-stage analysis the effect of environmental variables can by estimated by regressing the efficiency scores on these variables (Gerdtham, Rehnberg and Tambour, 1999; Kooreman, 1994; Luoma *et al.*, 1996). This is not a straightforward procedure though, since the efficiency scores are serially correlated and therefore the parameter estimates and the standard errors from these second-stage regressions are biased (Simar and Wilson, 2002). Also there is no consensual way of testing whether a particular environmental variable exerts a "significant" uncontrollable influence on efficiency and should therefore be controlled for (Dor, 1994).

In SFA, regressing the efficiency scores on environmental variables directly is not an appropriate procedure (Kumbhakar, Ghosh and McGuckin, 1991; and Reifschneider and Stevenson, 1991). In the first stage the inefficiency terms are assumed to be independently and *identically* distributed, which is inconsistent with assuming in the second stage they are a function of firm-specific factors. This inconsistency can be avoided by the use of alternative specifications (e.g. Battese and Coelli, 1995) that estimate efficiency and the effect of environmental variables on efficiency in a single stage procedure.

In comparison with DEA, decisions on the inclusion of environmental variables in SFA models are facilitated by the possibility of testing if these variables "explain" a material proportion of the overall residual[88]. In theory, including a new variable in the model should depend on whether it is thought to reflect an uncontrollable influence on the production frontier (in which case it would be included so that its effect is not mistakenly interpreted as (in)efficiency) or a characteristic of inefficient organisations (in which case it should be omitted from the production model and allowed to affect the estimates of efficiency[89]).

4.4.3.4 Dynamic effects

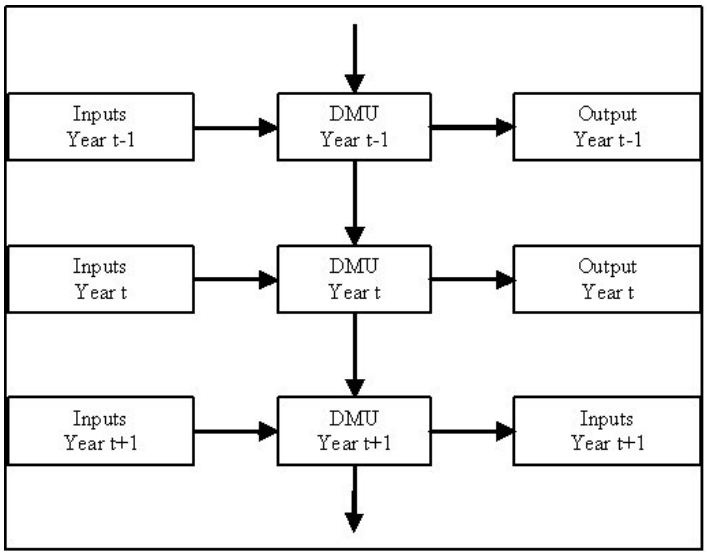
Organisations operate within historical contexts. They draw on past inheritances and make investments thinking in future performance[90]. Ideally, the production process should be modelled in a way that takes account of these dynamic effects - current performance being influenced by the endowment from previous organisational efforts, and contemporary inputs being partially invested for future output (Figure 4.8).

[88] The statistically significant effect of these variables may be due to one or both of the following reasons (Smith and Street, 2005): (a) it explains some of the random error (the original model suffered from omitted variable or functional form misspecification); (b) it explains some of the inefficiency error (the variable is correlated with the original estimates of inefficiency).

[89] Omitting the variable from the production frontier does not mean it should be omitted from the model altogether. It will still be important to test the effect of the variable on efficiency (in a second stage procedure in DEA, or in a single stage model in SFA). In DEA, this will provide information on the significance and magnitude of the effect the variable has on the efficiency. In SFA, the significance and magnitude of the effect will also be estimated, and in addition the efficiency estimates will be improved in case the effect of the new variable is significant.

[90] In other words, there are lags between inputs and outputs – e.g. increased public expenditure at a given moment in time may have a bearing on improved output indicators only years later. This seems to be particularly applicable to some government services such as in education and health (Atkinson, 2005).

Figure 4.8 : Dynamic effects



In practice, modelling the production process in this way is an intensely complex exercise. Difficulties arise in measuring endowments from previous periods; in ascertaining the impact some elements of current efforts will have on future attainment; and also in modelling the effects of these variables, together with the effects of contemporary inputs and outputs, on efficiency.

Existing examples of dynamic modeling of the production process include Färe and Grosskopf (1996) in DEA, and Bond (2002) in the parametric framework. However, even when adequate data is available, estimating these models often proves to be a deeply problematic exercise, impregnated with difficulties to which the literature is yet to provide a satisfactory answer.

4.5 Conclusion

Examples of public service efficiency estimation do not abound. Interesting work in this area includes the contributions of Coelli (1996a) for Australian Universities; De Borger and Kerstens (1996) for Belgian local governments; Fakin and Crombrugghe (1997); Evans, Tandon, Murray and Lauer (2000) for health expenditure; Gupta and Verhoeven (2001) for education and health in Africa; Johnes, Bradley and Millington (2001) for English secondary schools; Clements (2002) for education spending in Europe; Johnes, *et al.* (2002) for British higher education;

Greene (2004) for health expenditure; Afonso, Schuknecht and Tanzi (2005) for public expenditure in the OECD; Stevens (2005a; 2005b) for English and Welsh universities and for English local governments; Afonso and Aubyn (2004, 2005) for health and education expenditure in the OECD; and Afonso and Fernandes (2006) for local municipalities in the Lisbon Region.

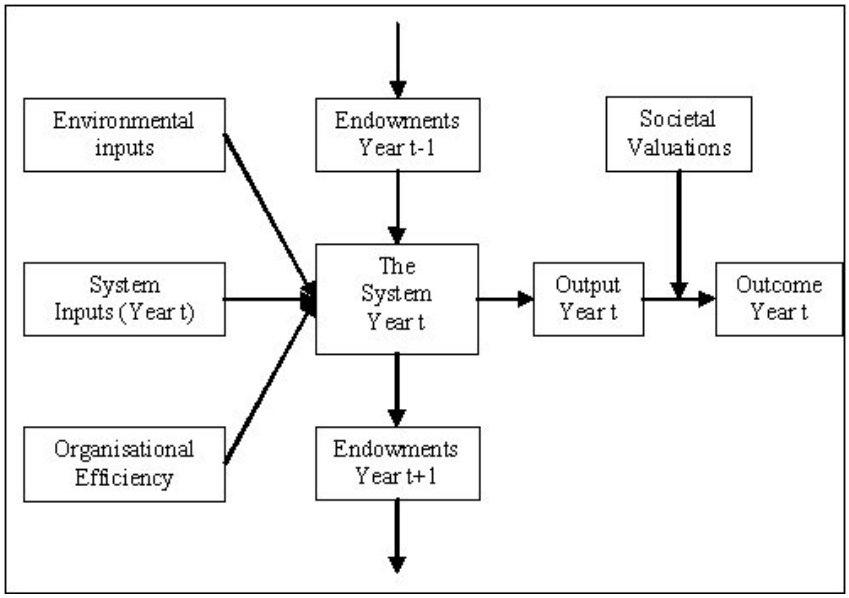
The most critical decisions in measuring efficiency involve the choice of the set of outputs to be measured and the choice of the set of weights reflecting the societal values attached to each output. In the public services, both of these are essentially political choices. Applying efficiency models to multiple-output technologies without an *a priori* definition of output weights imposes a set of assumptions on these weights[91], but it does not offer a solution to the political problem. This highlights the relevance of drawing on complementary approaches, such as trade-off surveys or conjoint analysis, which may enlighten the methodological choices required to measure efficiency (Shaw *et al.*, 2001; Ryan and Farrar, 2000).

The quality of the efficiency measures conveyed by the methods presented in this chapter will ultimately depend on the institutional framework under scrutiny. When a uniform set of objectives can be unequivocally applied to all organisations, the results are likely to be fairly robust. When organisations are thought to have an important degree of autonomy about the values they should pursue, the results may be less informative and more susceptible to criticism.

The quality of the results will also depend on the extent to which all the variables that significantly influence the productive process are taken into account, and their influence is correctly modelled. In its most general form, the production of public services should be modelled according to the structure depicted in Figure 4.9. However, in practice it is very unlikely that the information required for the estimation of such a comprehensive model can be obtained. Even if it could be obtained, modelling it would be of formidable complexity.

[91] DEA adjusts these weights so that the efficiency of each organisation is maximised, which is effectively equivalent to allowing each organisation to define its own set of weights. SFA (in the dual cost-function form) estimates these weights based on sample average costs, and thus assumes that the decisions made in practice concerning the distribution of expenditure between outputs effectively reflect societal valuations of those outputs.

Figure 4.9: Public service production model



Practicable empirical applications of efficiency models are necessarily infested with all sorts of compromises as they try to give parsimonious accounts of reality. Results are likely to be sensitive to these assumptions. So it is important to avoid relying on point estimates of efficiency for individual organisations. It is more sensible to simply sort the results into three groups: consistently efficient (may be indicative of best practice); consistently inefficient (may be indicative of sub-standard performance); and inconsistent results (avoid drawing strong inferences).

Chapter 5: Modes of Governance and Education Efficiency

5.1 Introduction

In section 7 of chapter 3 we laid out the main research hypothesis of this thesis. We argued that improving the efficiency of government activity implies that (i) governments retreat from interfering with transactions that do not exhibit attributes that call for public central coordination; (ii) when public central coordination is required, it should be confined to the specific attributes of transactions that markets fail to address, and in a way that explicitly acknowledges the risks of government failure. This chapter tests this hypothesis in the context of education services.

Applying our research hypothesis to the education sector yields specific expectations regarding the optimal organisational structure of education services from an efficiency point of view. Confining government interference to the attributes of education that call for public central coordination (externalities, market failures in general, and redistributive motives) provides for partial government involvement in the funding of education, but not necessarily in the provision of the service. It further suggests that when government directly provides the service through publicly owned and managed schools, these should be allowed to operate on an essentially decentralised manner, in line with the incentives that guide privately managed schools.

In the following sections we investigate the effect of organisational structure on efficiency. Efficiency is measured through a stochastic frontier for an unbalanced data panel of 18 OECD countries in 2000 and 2003. Organisational structure is explored through two analytical components: the share of public/private providers in the system and the degree of decentralisation of decision-making procedures of public providers.

The share of public providers is found to exert a negative effect on efficiency whereas the degree of decentralisation of the decision-making procedures of public providers is found to exert a positive effect on efficiency. Both results corroborate our research hypothesis.

The next section explores the attributes of education that call for public central coordination. Section 5.3 presents the variables used for measuring efficiency and the organisational variables whose effect on efficiency is being investigated. Section 5.4 focuses on the stochastic model that underpins the results presented in section 5.5. Section 5.6 concludes the chapter.

5.2 The Economics of Education

Since the early writings of Sir William Petty and Adam Smith, economists have come to regard education as an investment made by students and by society at large, each of these groups incurring their own costs and reaping specific rewards[92].

Students bear the costs of education in the form of direct costs of tuition and income forgone for attending school instead of taking a paying job. The benefits accrue in the form of higher average income in the course of the life-cycle, greater job opportunities, and less probability of becoming unemployed (OECD, 2001). In addition, several non-financial rewards reinforce the benefits obtained through the labour market. There is evidence of more educated people making better choices concerning health, thereby extending their life expectancy and improving their quality of life. There is also evidence of children of more educated parents performing better at school, leaving school later and learning more (Michael, 1982; Haveman and Wolfe, 1984; Wolfe and Zuvekas, 1995; and Leibowitz, 1974).

Society as a whole incurs the direct costs of providing free or heavily subsidized education to its citizens, and forgoes the opportunity to devote the resources engaged in education to other projects. The benefits stemming from a well educated citizenry come in various forms. National income increases directly with the higher payoffs that educated workers obtain in the labour market. Civic involvement and electorate responsibility are likely to improve with education (Teixeira, 1992). The level of crime is likely to decrease (Ehrlich, 1975; Lochner and Moretti, 2001). Furthermore, some studies suggest that education provides economic benefits to society that exceed the sum of its benefits to individuals (i.e. education generates external effects), by contributing to the creation of an appropriate environment for innovation, and thus accelerating the growth rate of the economy (Lucas, 1988; Romer, 1990; Barro, 1991; Jorgenson and Fraumeni, 1992; Barro and Sala-I-Martin, 1995).

Although associations between education and various desirable economic outcomes are well established, significant questions remain about the magnitude and interpretation of these relationships (Hanushek, 2002). First, the direction of causality

[92] Schultz (1961; 1963), Becker (1993) and Mincer (1970, 1974) were responsible for the introduction of the conceptualisation of education as an investment into mainstream economics, and for creating the basis for a steady stream of subsequent theoretical and empirical analyses (Hanushek, 2002).

intrinsic to these relationships is still the subject of controversy (on the link between education and economic growth see Bils and Klenow, 2000; Mankiw, Romer, and Weil, 1992; and Benhabib and Spiegel, 1994). Second, average effects and marginal effects may not coincide, which further complicates the analysis. Third, the external effects often associated with education are elusive and difficult to estimate convincingly.

The existence of large returns to education is not by itself a reason for government to interfere with the production and/or the consumption of education. It is simply a powerful incentive for individuals to invest in human capital accumulation. Government involvement in the education sector is usually justified on the basis of externalities, market failures in general, and redistributive motives (Hanushek, 2002).

Despite being elusive and difficult to estimate, there appears to be a consensual sense that education involves substantial externalities (Cohn and Geske, 1990). The main candidates for the role of positive external effect generators are citizen involvement in the community and government, crime reduction, family decision making and child upbringing, and economic growth. Public central coordination can allow for these external effects and bring individuals decisions in line with the appropriate social calculus.

Other types of market failures that may lead government to intervene include capital market and information imperfections. If individuals cannot barrow against their human capital, they may be forced to underinvest in education (Becker, 1993). This possibility has not received strong empirical support (Cameron and Heckman, 1999), but the current level of government interference in primary and secondary education hinders the proper evaluation of the real significance of this problem. Also, if informational problems prevent individuals from accurately assessing the influence of schools on their performance[93], they may be unable to make optimal human capital investment decisions. In both instances public central coordination can remove/mitigate the hindrances to efficient individual decision-making.

The remaining argument for government interference in education is usually entrenched in the desire to change the prevailing distribution of income (Juhn, Murphy and Pierce, 1993). The precise definition of societal redistributive goals, and the way they interact with education remain largely unexplored.

[93] Apart from school inputs, student performance is generally affected by a mixture of student innate ability and a range of environmental factors (namely the family socio-economic background).

While these attributes can be legitimately invoked to justify government involvement in education, none of them provides for extensive public funding of education or direct public provision of education services.

Externalities and capital market imperfections lead to sub-optimal investment on education. Public funding is therefore required to motivate individuals to investment more on education, but this is likely to be residual. Using the education system as a redistributive mechanism also implies that a certain (residual) share of the system be publicly funded. Information imperfections, on the other hand, bring a role for government in terms of facilitating information, but have no implications for the funding or provision of the service.

In virtually all the OECD countries, the organisation of primary and secondary education has moved toward extensive government financing and government provision. Our research hypothesis refutes the adequacy of such a large intervention. In the next sections we place our hypothesis under empirical scrutiny.

5.3 Education Efficiency and Organisational Structure

There are a limited number of studies addressing the efficiency of education systems at the international level (Clements, 1999, 2002; St. Aubyn, 2002; Gupta and Verhoeven, 2001; Afonso and St. Aubyn, 2004, 2005). These studies follow a parallel line of research to the traditional econometric estimation of education production functions using cross-country data (Barro and Lee, 2001; Hanushek and Kimko, 2000; Hanushek and Luque, 2003). In both approaches, socio-economic variables describing students' backgrounds such as parents' education and wealth are the key variables used to explain efficiency[94]. Whilst the importance of these variables cannot be overstated, they are of little help to policymakers attempting to improve the outcomes of education expenditure since (at least in the sort run) they are beyond government control. More interesting from a policy point of view is the effect that organisational

[94] The literature focusing on the econometric estimation of education production functions does not explicitly acknowledge the concept of *efficiency*. Typically, qualitative measures of education outputs (results from international studies assessing students' academic abilities) are regressed on a set of education system resources (financial and/or physical) and other environmental variables thought to affect the performance of the education system. However, since students' socio-economic background (normally proxied by parents' education and wealth) is often presented as a relevant environmental variable for explaining students' performance, it is implicitly assumed that it interferes with the relationship between inputs and outputs, and thus with efficiency.

structure of education systems has on efficiency (something that government act upon). The nature and significance of this effect is explored in the rest of this chapter.

The traditional conceptualisation of the education production process is of the following form (Hanushek and Taylor, 1990):

$$A = f(S\ ;F\ ;\varepsilon\ )$$

(1)

A: a vector of variables measuring individual achievement;

S: a vector of variables measuring school inputs;

F: a vector of variables measuring students' socio-economic background;

ε: a vector of unmeasured factors that contribute to individual achievement (individual ability and random factors).

Here, we expand this model to $A = f(S\ ;F\ ;O;\varepsilon\ )$, where O is a vector of variables measuring organisational structure.

For measuring student achievement we draw on the results of two OECD surveys conducted in 2000 and 2003 which evaluate the mathematical, reading and scientific ability of 15 year old students[95].

Several different variables have been considered in the education efficiency literature to measure the school inputs contributing to student achievement. These have typically included financial indicators such as expenditure per student and physical indicators such as average class size, ratio of students to teaching staff, and number of instruction hours. Financial inputs were predominant in most of the pioneer studies on education efficiency. Recent studies have favoured the use of multiple physical inputs instead - often being alleged that financial inputs reflect to a large extent differences in costs between countries (mainly teachers' salaries) causing significant distortions to the measurement of efficiency.

[95] An arithmetic average of the mathematic, reading and science scores from the 2000 and 2003 *Programme for International Student Assessment (PISA)* surveys was considered. Since the reporting scales for mathematics are not directly comparable between 2000 and 2003, as the PISA 2003 mathematics assessment was more comprehensive than the one that took place in 2000, we drew on the two components of the 2003 mathematical results (*space and shape* and *change and relationships*) common to the 2000 survey.

We employ a single financial indicator - cumulative expenditure per student[96] (CEPS) between the age of 6 and 15 in equivalent US dollars - as the input variable for three different reasons. First, differences in costs underlying financial measures of inputs can be mitigated by the use of values converted at purchasing power parities (PPPs) - although they are not totally eliminated as PPPs are calculated taking into account the average level of prices (the one that underlies the gross domestic product) and not the particular prices/costs of specific services like education. Second, some of these differences in cost actually reflect differences in the overall efficiency of the system that should be considered by our calculations. For example, the skill and ability of school managers to contract the best teachers at the best price, to put in place an incentive regime that ensures high labour productivity, or to outsource non-core activities in a cost effective manner, create differences in costs that should be reflected in a proper account of efficiency. Third, despite all the imperfections and flaws of international data on education expenditure, this still arguably remains the best available proxy for the entire set of resources that go into the production of education. Resources such as physical infrastructure (buildings, computers, etc) and policy administration (of schools and the education system as a whole) escape traditional physical input measures, but have a bearing on financial inputs.

Students´ socio-economic background is proxied by the PISA's index of economic, social and cultural status (IESCS). This indicator captures various aspects of students' family and home background that are thought to be relevant for school achievement[97].

Organisational structure is expressed by two variables: the share of public providers in the system (%PUBPROV), and the degree of decentralisation/managerial

[96] Let *n(0), n(1)* and *n(2)* be the typical number of years spent by a student from the age of six up to the age of 15 years in primary, lower secondary and upper secondary education. Let *E(0), E(1)* and *E(2)* be the annual expenditure per student in US dollars converted using purchasing power parities in primary, lower secondary and upper secondary education, respectively. The cumulative expenditure is then calculated by multiplying current annual expenditure *E* by the typical duration of study *n* for each level of education *i* using the following formula, (OECD, 2004b):

$$CE = \sum_{i=0}^{2} n(i) * E(i)$$

[97] IESCS is derived from the following variables (OECD, 2004b, p. 307): *i)* the highest international socio-economic index of occupational status of the father or mother; *ii)* the highest level of education of the father or mother converted into years of schooling; and *iii)* the number of books at home as well as access to home educational and cultural resources, obtained by asking students whether they had at their home: a desk to study at, a room of their own, a quiet place to study, a computer they can use for school work, educational software, a link to the Internet, their own calculator, classic literature, books of poetry, works of art (e.g., paintings), books to help with their school work, and a dictionary.

autonomy of public providers (DEC). In the following paragraphs we provide a brief discussion of the rationale for including these variables.

Primary and secondary education is mainly a public enterprise. On average across OECD countries only 4 per cent of 15-year-olds are enrolled in schools that are both privately managed and predominantly privately financed (OECD, 2004b). However, publicly financed schools are not necessarily publicly managed, and in fact, in the quest for efficiency gains, governments are increasingly relying on the transfer of public funds to privately managed schools to deliver education services. Accordingly, schools that are privately managed but predominantly publicly financed[98] account for a more expressive average of 13 per cent of school enrolment in OECD countries.

In contrast to previous studies that have tested (unsuccessfully) the effect of the ratio of public-to-total expenditure on efficiency[99], we argue that the share of public providers (schools both funded and managed by the public sector) offers a better account of the structural diversity of education systems and that its effect on efficiency is the one that should be investigated.

Apart from the segmentation between public and private providers, a second organisational feature of education systems whose impact on efficiency is worth studying relates to the reforms that have been taking place throughout the last two decades in terms of the way the public sector operates as a service provider. In particular, while seeking to improve efficiency, many governments have opted for devolving decision-making authority to schools and lower levels of governments. This has been accomplished in the expectation it will enable downsizing of central education administration; elimination of superfluous layers of bureaucracy; and improvement of chains of command in decision making, delivering a larger proportion of financial and human resources directly to local governments, schools, and students (Behrman, *et al.*, 2002).

---

[98] OECD (2004b) designates these as "government-dependent schools". Public schools are defined as educational instructional institutions that are controlled and managed directly by a public education authority or agency; or controlled and managed either by a government agency directly or by a governing body (council, committee, etc.), most of whose members were either appointed by a public authority or elected by public franchise. Private schools are defined as educational instructional institutions that are controlled and managed by a nongovernmental organisation (*e.g.,* a church, a trade union or a business enterprise) or if their governing board consisted mostly of members not selected by a public agency.

[99] Afonso and St. Aubyn (2005) did not find a significant relationship between public-to-total education expenditure and efficiency. The authors argued that this was "probably because most spending in this level of education is essentially public and high for most countries" (p. 23).

We examine the effect of decentralisation/managerial autonomy on efficiency using an indicator of the distribution of educational decisions by specific levels of government in public lower secondary education (OECD, 2004a). The percentage of decisions taken at the "local" and "school" levels is taken as the appropriate indicator of decentralisation/managerial autonomy. Information on four different functional domains is considered:

- *Organisation of instruction*: student admissions; student careers; instruction time; choice of textbooks; grouping students; additional support for students; teaching methods; regular day-to-day student assessment;

- *Personnel management*: hiring and dismissal of teaching and non-teaching staff; duties and service conditions of staff; salary scales of staff; influence over the careers of staff;

- *Planning and structures*: opening or closure of schools; creation or abolition of a grade level; design of programmes of study; selection of programmes of study taught in a particular school; choice of range of subjects taught in a particular school; definition of course content; setting of qualifying examinations for a certificate or diploma; credentialing (examination content, marking and administration);

- *Resources*: allocation and use of resources for teaching staff, non-teaching staff, capital and operating expenditure.

Our model is estimated for an unbalanced panel of 18 OECD countries in 2000 and 2003[100]. Data for all variables are presented in the statistical annex and descriptive statistics are presented in Table 5.1.

Table 5.1: Descriptive Statistics

Variable	Mean	Std. Dev.	Min.	Max.
PISA scores	494.000	30.811	393.000	544.000
Cumulative expenditure per std	44148.406	18548.700	10480.545	73328.892
IESCS	-0.020	0.359	-1.130	0.690
%PUBPROV	0.846	0.187	0.233	0.995
Dec	0.603	0.261	0.156	1

[100] Data on cumulative expenditure refers to 1998 and 2002 (deflated to 1998 prices). Data on decentralisation of public providers also refers to 1998 and 2003.

5.4 Methodology

We use the Battese and Coelli (1995) model for a stochastic frontier production function which is equivalent to the Kumbhakar, Ghosh and McGukin (1991) specification, with the exception that allocative efficiency is imposed, the first-order profit maximising conditions removed, and panel data is permitted. The model may be expressed as:

$$Y_{it} = X_{it}\beta + (V_{it} - U_{it}) \qquad i=1,...,18;\ t=1,2 \qquad (2)$$

Y_{it} is the logarithm of the PISA score of country i in period t;

X_{it} is the logarithm of cumulative expenditure per student of country i in period t;

β is a vector of unknown parameters;

t denotes the time period;

V_{it} are random variables assumed to be independent and identically distributed $N(0,\sigma_V^2)$, and independent of U_{it};

U_{it} are non-negative random variables assumed to account for technical inefficiency in production and assumed to be independently distributed as truncations at zero of the $N(m_{it},\sigma_U^2)$ distribution, where:

$$m_{it} = Z_{it}\delta, \qquad (3)$$

Z_{it} is a $p\times1$ vector of variables deemed to influence the efficiency of country i in period t; δ is an $1\times p$ vector of parameters to be estimated.

Following Battese and Corra (1977), the likelihood function is parameterised in terms of the variance ratio $\gamma=\sigma_U^2/(\sigma_V^2+\sigma_U^2)$. Hence γ indicates the relative magnitude of technical inefficiency variance to total variance in the model.

Six different specifications of equation (3) are tested:

Model a: $m_{it} = \delta_1 IESCS_{it}$;

Model b: $m_{it} = \delta_0 + \delta_1 IESCS_{it} + \delta_2 \ln(\%PUBPROV)$;

Model c: $m_{it} = \delta_1 IESCS_{it} + \delta_2 \ln(DEC)$;

Model d: $m_{it} = \delta_1 IESCS_{it} + \delta_2 \ln(\%PUBPROV) + \delta_3 \ln(DEC)$;

Model e: $m_{it} = \delta_1 IESCS_{it} + \delta_2 \ln(\%PUBPROV) + \delta_3 \ln(\%PUBPROV*DEC)$;

Model f[101]: $m_{it} = \delta_0 + \delta_1 IESCS_{it} + \delta_2 \ln(DECINDEX)$

Models "a", "b", and "c" test the isolated effect on efficiency of IESCS, %PUBPROV and DEC, respectively[102]. Models "d", "e" and "f" include simultaneously the three variables and consider different specifications for %PUBPROV and DEC. Model "d" treats the effects of %PUBPROV and DEC on the general efficiency of education systems as being autonomous and independent from each other. The implicit assumption is that decentralisation of decision making in the public sector has an impact on the efficiency of the education system as a whole that is independent of the public share of the system (i.e. the effects of decentralisation in the public sector spillover to other parts of the system – possibly through competition with private and not-for-profit providers). Model "e" considers that the effect of decentralisation on efficiency is mediated by the share of public providers in the education system (i.e. there are no spillover effects). Finally, model "f" tests the effect on efficiency of an overall index of decentralisation that assumes private schools correspond to a level of decentralisation of one and that public schools are in this respect comparable to private schools, i.e. the type of ownership is irrelevant. In the different specifications an independent term was included when its estimate was significant[103].

5.5 Results

Maximum likelihood estimates of the parameters in the production function frontier and technical inefficiency effects for the different specifications are presented

[101] DECINDEX= 1-[%PUBPROV(1-DEC)]. This index assumes that private schools correspond to a level of decentralisation of 100 per cent. The index is equivalent to an average of the level of decentralisation of public and private schools, weighted by the share of public and private providers in the education system, respectively.

[102] In these two later cases controlling for IESCS.

[103] We performed generalized likelihood ratio tests to determine whether or not an independent term should be included.

in Table 5.2. The efficiency ranking associated with model "e"[104] is presented in Table 5.3.

Table 5.2: Maximum Likelihood Estimation Results. Absolute t-ratios in Parentheses

Model a					
Production function		Technical inefficiency effects		Variance parameters	
Parameter	Estimate	Parameter	Estimate	Parameter	Estimate
Constant	6.100	IESCS	-0.187	σ^2	0.002
	(30.787)		(-4.253)		(2.813)
CEPS	0.012	Log Likelihood	52.274	γ	0.125
	(0.686)				3.410
Model b					
Production function		Technical inefficiency effects		Variance parameters	
Parameter	Estimate	Parameter	Estimate	Parameter	Estimate
Constant	6.205	Constant	0.047	σ^2	0.002
	(3.871)		(1.887)		(4.132)
CEPS	0.004	IESCS	-0.180	γ	0.196
	(0.257)		(-4.538)		(1.183)
		ln(%PUBPROV)	0.117		
			(1.997)		
		Log Likelihood	56.366		
Model c					
Production function		Technical inefficiency effects		Variance parameters	
Parameter	Estimate	Parameter	Estimate	Parameter	Estimate
Constant	6.114	IESCS	-0.143	σ^2	0.002
	(34.016)		(-2.856)		(3.215)
CEPS	0.012	ln(DEC)	-0.031	γ	0.258
	(0.724)		(-1.295)		(0.828)
		Log Likelihood	53.113		
Model d					
Production function		Technical inefficiency effects		Variance parameters	
Parameter	Estimate	Parameter	Estimate	Parameter	Estimate
Constant	6.131	IESCS	-0.140	σ^2	0.002
	(50.426)		(-3.185)		(4.521)
CEPS	0.011	ln(%PUBPROV)	0.135	γ	0.391
	(0.940)		(2.469)		(1.543)
		ln(DEC)	-0.055		
			(-3.150)		
		Log Likelihood	57.667		
Model e					

[104] The one that generated the highest log likelihood value.

Production function		Technical inefficiency effects		Variance parameters	
Parameter	Estimate	Parameter	Estimate	Parameter	Estimate
Constant	6.128	IESCS	-0.139	σ^2	0.002
	(42.047)		(-3.152)		(2.187)
CEPS	0.011	ln(%PUBPROV)	0.176	γ	0.344
	(0.809)		(4.014)		(0.923)
		ln(%PUBPROV*DEC)	-0.055		
			(-2.816)		
		Log Likelihood	57.780		

Model f

Production function		Technical inefficiency effects		Variance parameters	
Parameter	Estimate	Parameter	Estimate	Parameter	Estimate
Constant	6.109	Constant	-0.083	σ^2	0.002
	(32.700)		(-1.615)		(1.144)
CEPS	0.012	IESCS	-0.166	γ	0.248
	(0.677)		(-3.150)		(0.316)
		ln(DECINDEX)	-0.105		
			(-2.349)		
		Log Likelihood	56.211		

Table 5.3: Efficiency Ranking

Country	Output Efficiency	Country	Output Efficiency
2000		2003	
Finland	0.991	Netherlands	0.997
Norway	0.990	Korea	0.995
Denmark	0.990	Finland	0.994
Korea	0.989	Iceland	0.993
Hungary	0.976	Norway	0.991
Austria	0.974	Czech Republic	0.990
Czech Republic	0.973	Denmark	0.989
Germany	0.972	Sweden	0.989
Spain	0.952	Japan	0.989
Italy	0.925	Hungary	0.980
Greece	0.909	Germany	0.978
Portugal	0.872	Slovak Republic	0.971
		Austria	0.971
		Italy	0.948
		Spain	0.942
		Portugal	0.890
		Greece	0.893
Average Efficiency	0.960	Mexico	0.792

The results are consistent between the different models. Apart from the expected positive effect of students' socio-economic background on efficiency (reflected in the negative estimate of the technical inefficiency coefficient), the share

of public providers is found to exert a negative effect on efficiency whereas decentralisation is positively associated with efficiency (the technical inefficiency parameters are positive and negative, respectively).

The extreme cases of (in)efficiency presented in Table 5.3 provide good illustrations of these results. In general, the most efficient countries (Finland, Norway, Denmark and Korea in 2000; the Netherlands, Korea, Finland, Iceland, Norway, C.Republic, Denmark, Sweden and Japan in 2003) have better economic and cultural status, higher levels of decentralisation of decision making in the public sector, and lower shares of public providers[105] in comparison with the most inefficient countries (e.g. Italy, Greece, and Portugal in both 2000 and 2003; and Spain and Mexico in 2003).

We tested the robustness of these results by performing a series of generalized likelihood ratio tests. The likelihood ratio statistic is given by $\lambda = 2(\ln L_1 - \ln L_0)$, where ln L_0 and ln L_1 are the maximum log-likelihood values under the null and alternative hypotheses, H_0 and H_1, respectively. Under the null hypothesis this statistic is usually assumed to be asymptotically distributed as a chi-square random variable with degrees of freedom equal to the number of restrictions involved in the test. However, when the null hypothesis involves a restriction of the type $\gamma=0$ this statistic can be shown to have an asymptotic non-standard mixed qui-square distribution with degrees of freedom equal to the number of parameters restricted to zero under the null hypothesis (Coelli *et. al*, 1998)[106]. The results of the likelihood ratio tests are presented in Table 5.4.

Table 5.4: Likelihood Ratio Tests (5% level of significance)

Model b			
Null hypothesis	λ	Critical Value	Decision
H_{01}: ln(%PUBPROV)=0	8.367	3.84	Reject
H_{02}: λ=0, Constant=0; IESCS=0; ln(%PUBPROV)=0	26.507	8.761	Reject
Model c			
Null hypothesis	λ	Critical Value	Decision
H_{01}: ln(DEC)=0	1.678	3.84	Not Reject
H_{02}: λ=0; IESCS=0; ln(DEC)=0	20.001	7.045	Reject
Model d			

[105] Except for Iceland (2003), Norway (2000 and 2003), Finland (2003) and Sweden (2003), where the share of public providers is relatively high.
[106] We extracted the critical value for this statistic from Kodde and Palm (1986).

Null hypothesis	λ	Critical Value	Decision
H_{01}: ln(%PUBPROV)=0	9.107	3.84	Reject
H_{02}: ln(DEC)=0	7.338	3.84	Reject
H_{03}: ln(%PUBPROV)=0; ln(DEC)=0	10.785	5.99	Reject
H_{04}: λ=0; IESCS=0; ln(%PUBPROV)=0; ln(DEC)=0	29.109	8.761	Reject
Model e			
Null hypothesis	λ	Critical Value	Decision
H_{01}: ln(%PUBPROV)=0	10.967	3.84	Reject
H_{02}: ln(%PUBPROV*DEC)=0	7.564	3.84	Reject
H_{03}: ln(%PUBPROV)=0; ln(%PUBPROV*DEC)=0	11.011	5.99	Reject
H_{04}: λ=0; IESCS=0; ln(%PUBPROV)=0; ln(%PUBPROV*DEC)=0	29.335	8.761	Reject
Model f			
Null hypothesis	λ	Critical Value	Decision
H_{01}: ln(INDEXDEC)=0	8.057	3.84	Reject
H_{02}: λ=0, Constant=0; IESCS=0; ln(INDEXDEC)=0	26.197	8.761	Reject

The first set of null hypotheses in each model tests the individual and combined significance of the organisational variables DEC and %PUBPROV. The last null hypothesis tests the appropriateness of the stochastic frontier and technical inefficiency effects specification by assessing the possibility of it being equivalent to the average response function, which can be efficiently estimated by ordinary least squares regression. These null hypotheses are clearly rejected[107] and so we conclude that the effects of our organisational variables are significant and that the stochastic specification of the efficiency frontier is appropriate.

5.6 Conclusion

The education sector is undergoing intense reforms. New modes of delivery and new provider structures, mostly drawn from private sector practice, are being tested and adapted to the political and institutional context in which education services operate. Governments are increasingly relying on private providers as a means of defining institutional environments that induce efficient behaviour. At the same time, the public sector itself is redefining the way it operates as a service provider – from a traditional purely bureaucratic form of organisation to the constitution of effective service agencies, highly decentralised, and operating along

[107] Except for the individual effect of DEC in model "c". However, since we expect this effect to be influenced by %PUBPROV this lack of individual significance does not appear to be particularly meaningful.

privately-inspired management criteria. Overall, these tendencies are in line with the research hypothesis laid out in the previous chapter. The empirical results presented in the current chapter corroborate our hypothesis that these practices should have positive effects on efficiency.

Chapter 6: Modes of Governance and Health Efficiency

6.1 Introduction

This chapter investigates the research hypothesis formulated in the seventh section of chapter 3 within the context of health care services. According to this hypothesis, the optimal organisational structure in terms of efficiency is one where governments confine their interference to the specific attributes of transactions that markets fail to address, and in a way that allows for risks of government failure.

Governments can be involved in the health sector in various ways, the most significant of which is in guaranteeing that the entire population is protected against the financial risks of sickness and medical treatment. The second most significant role is in the provision of medical services, by owning and operating medical care providers. In addition to intervening in the funding and provision of health services, governments can tax goods with adverse effects on health, and regulate the health sector – defining the operational framework of insurance companies; issuing licenses for medical care providers; and (dis)approving the commercialisation of new drugs and devices.

Our research hypothesis raises specific expectations as to the optimal form of government involvement from an efficiency point of view. Contrary to the education sector, confining government interference to the attributes of health transactions that call for public central coordination (mainly information imperfections) provides for extensive government involvement in health funding/insurance, and simultaneously significant involvement in the provision of services. On the other hand, and as in the case of education, it also suggests that public health care providers should be given sufficient autonomy to explore the benefits of decentralised coordination, and should be made to operate in an environment open to competitive forces.

Several OECD countries have recently introduced substantial reforms to the organisational structure of their health services, in an attempt to stimulate innovation and flexibility in the way health systems respond to political priorities. These reforms have included experimenting with market-inspired mechanisms in the internal operation of public health care providers and in their interaction with private providers. This chapter draws on the reform experience of 23 OECD countries to test the effect of health modes of governance on efficiency, and assess the empirical validity of our research hypothesis. The results show that the introduction of market-type mechanisms to public integrated health systems has a positive effect on

efficiency, whereas further movement towards a market model of health care insurance and provision (public contract, and private insurance/provider models) reduces efficiency.

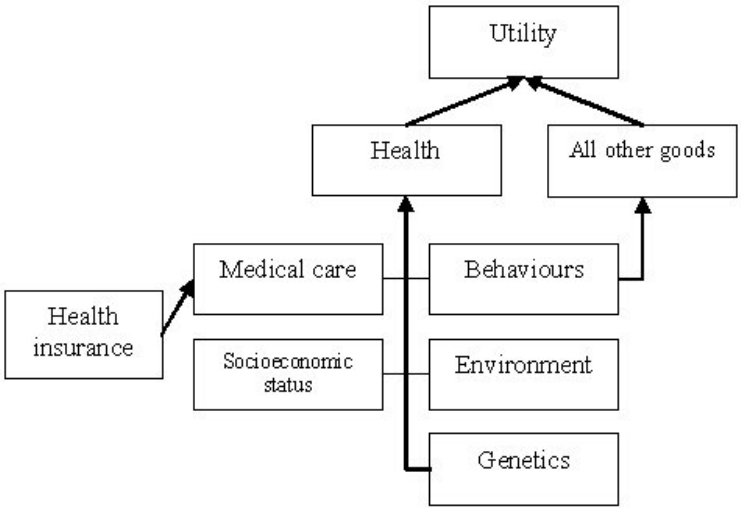
The chapter proceeds as follows. Section two examines the attributes of health that call for public central coordination. Section three presents various modes of organising transactions in the health sector and discusses their expected effect on efficiency in the light of our research hypothesis. Section four focuses on the methodological issues that underlie the international comparative measurement of health service efficiency. Section five presents DEA and SFA efficiency estimates for four different input-output datasets. Section five concludes the chapter.

6.2 The Economics of Health

The importance of the health sector can hardly be overstressed. Health (like education) is among the basic capabilities that give value to human life. The opportunity to receive basic education or health care, together with freedom of political participation, form a set of substantive freedoms which are simultaneously constitutive components of human development (end goals) as well as drivers of economic progress (Sen, 1999).

In order to understand the attributes of transactions in the health sector whose optimal coordination requires government interference, we need to look at the variables that influence the health status of individuals and their well being (Figure 6.1). Individual utility is a function of health status and consumption of all other kinds of goods. Health status, in turn, depends on individual behaviours (which also affect utility directly); environmental factors; medical care; and other factors such as genetics and socioeconomic variables.

Figure 6.1: Health and utility (adapted from Cutler, 2002, p. 2147)



Health-related behaviours such as smoking and drinking yield benefits for individuals (direct consumption value), but they also bring negative health consequences both to the individuals who adopt these behaviour and possibly others. Similarly, health-related environmental factors such as pollution generate benefits for firms and consumers (if pollution is a by-product of industrial activity), but they also have harmful side-effects for health. In both instances, government intervention is required to correct these external effects and induce individual (and managerial) decisions to allow for the full societal costs of these behaviours.

In addition to externalities, some health-related behaviours have merit good attributes which may also justify government intervention. In particular, if individual decisions on the consumption of goods or services harmful for health are short-sighted, mostly driven by impulse and/or addiction, taxing or banning consumption may be necessary to induce people to internalise the full consequences of their behaviour.

Individual health status is in itself a source of external effects. From a negative perspective, individual sickness and sickness-related behaviours may have direct negative consequences for other members of society (e.g. contagious diseases; development of antibiotic-resistant bacteria from inappropriate use of antibiotics). From a positive perspective, health is (together with education) a cornerstone of human capital (Becker, 1975), which in turn plays a key role in economic growth (Bleaney, Kneller and Gemmel, 2001; Bloom, Canning and Sevilla, 2001). Since individual decision-making is unlikely to reflect these externalities, public central

coordination is required to influence the consumption of health care and/or regulate health-related behaviours.

The most important role for government in the health sector, however, is in the market for health care and its subsidiary health insurance. Medical care and insurance markets are plagued with informational problems: moral hazard and adverse selection in health insurance; incomplete information on the part of patients in health care; asymmetric information between consumers and producers about patients needs; and inability to determine the quality of services, even after they have been provided (Cutler, 2002). These informational problems lead free competitive markets for medical care and insurance to failure[108].

Health insurance[109] markets are prone to moral hazard problems – people are likely to consume more health services when insured than they would do if they had to pay the full price of these services. In addition, health insurance markets are also known to suffer from adverse selection problems. Assuming individuals represent heterogeneous medical risks, and insurers are able to identify individual risk types and price policies accordingly, a free insurance market will ensure that individuals are fully insured, but higher risk people (who seek more generous health insurance) will pay higher premiums. This equilibrium does not pool health risks, and is therefore inefficient. If it is assumed instead that knowledge of individual risk type is limited or insurers are not allowed to use such knowledge in setting prices, high risk individuals will tend to drive up the price of more generous plans, while the other plans remain relatively cheap. This situation is also inefficient, and it gives rise to three sources of welfare loss: first, a risk segmentation loss, as sick individuals still pay more for insurance than the healthy; second, given the differences in price, marginal people are inappropriately induced to buy less generous insurance plans in an attempt to benefit from lower premiums; third, insurance plans tend to be less generous in their benefits so as to attract the healthy and repel the sick. In either case public central coordination

[108] Apart from externalities, merit good attributes, and information imperfections, equity is another important motivation for government involvement in medical care. Modern societies have come to look at medical treatment not just as an another ordinary service but has a right of citizenship, regardless of the financial position of the individual in need, or their past efforts to provide for health hazards. Ensuring universal access to medical services is often an important driver of government involvement in health services.

[109] The goal of health insurance is to equalise the marginal utility of income in different states of nature, which in many cases is achieved by smoothing the financial risk associated with medical costs.

is required so that some form of mandatory, universal insurance plan may be enforced[110].

In medical care, informational problems also hinder the coordinative efficiency of free competitive markets. First, information asymmetries between patients and physicians as to the complexities of medical care diagnosis and treatment give market power to the later. Second, medical care has attributes of a credence good, i.e. ascertaining the quality of the service is difficult, even after it has been provided (Tirole, 1988; Darby and Karni, 1973). The combination of profit-maximisation with these two forms of information imperfections contributes to disseminating perverse incentives that lead competitive markets to failure. Private for-profit suppliers will have an incentive to skimp on quality if consumers and/or government lack the regulatory and monitoring ability to detect such behaviour.

The coexistence of public and not-for-profit together with for-profit medical care providers may be seen as a means of counteracting the perverse incentives that pure profit maximisation would lead to (Arrow, 1963), as these different organisational forms are expected to influence each other in the market place[111]. It is also a way of avoiding the perverse incentives that could be created if the public sector held the monopoly of medical care service provision – lack o competitive pressure for innovation and efficiency, and the typical shortcomings of public bureaucracies (addressed in section 4.2, chapter 3).

6.3 Modes of Governance and Health Efficiency

Health services can be organised in three fundamentally different ways – via hierarchical bureaucracies, through long-term contractual arrangements under some

[110] In practice, all OECD countries have some form of publicly financed or administered health insurance programmes. Private health insurance is the dominant form of basic coverage in the United States and Switzerland, and covers a sizeable minority of the population in Germany and the Netherlands. In countries such as Hungary, Japan, Luxembourg, Mexico, and most Nordic countries, private health insurance policies are not commonly used. In other countries, private health insurance is used to fill gaps in the benefits package (a supplemental policy) or absorb out-of-pocket payments (complementary insurance). Private insurance duplicates coverage provided by universal public programmes in Australia, Ireland, Italy, Spain, and the United Kingdom, where such coverage is purchased mainly to increase choice of providers and timeliness of care (Docteur and Oxley, 2003).

[111] Hansmann (1980) argued that the quality at for-profit hospitals in the US was kept high because not for-profit competitors provided high quality services, making deviations from quality by for-profit hospitals more readily detectable. However, Cutler and Horwitz (1999), and Silvermann and Skinner, (2000) drew attention to the inverse effect, i.e. for-profit hospitals may lead not-for-profit hospitals to change their behaviour in socially adverse ways.

degree of non-market control[112], and as direct, short-term market-based interactions between patients and providers. These arrangements are independent of whether ownership is public or private[113] (WHO, 2000).

The way health services are financed plays a critical role in the design of their optimal mode of governance. An adequately financed health system should ensure universal financial protection. It should avoid exposing individuals to large unexpected expenses, relying as fully as possible on predictable prepayment (insurance) and minimising the share of out-of-pocket funding. It should also effectively pool the financial risks of healthy and sick individuals. Problems of moral hazard and adverse selection (discussed in the previous section) prevent free competitive insurance markets from achieving these goals efficiently, which prompts extensive government intervention in the funding of health care[114]. This result motivates our first hypothesis regarding the effect of health organisational structure on efficiency: modes of governance based on direct, short-term market interactions between patients and providers, or relying mainly on unregulated private insurance are expected to be relatively inefficient.

At the service provision level, market modes of governance are generally perceived as being more effective in stimulating innovation and flexibility in the way the health system responds to specific needs. However, the combination of profit-maximisation with imperfect information propagates perverse incentives that lead competitive markets to failure. Public hierarchical modes of governance have the

[112] One type comprises long-term contracts between the public sector and nongovernmental providers (both non-profit and for-profit). Another contractual arrangement characterizes private insurance, which may or may not be publicly regulated.

[113] WHO (2000, p. 63) provides an interesting example: "ownership of services that are organized as hierarchies can be public, as in the extensive network of public health, hospital, and ambulatory clinics that are part of the Turkish Ministry of Health service delivery system and that of many other countries. But they can also be private, as in a United States health management organisation like Kaiser Permanente. Such private entities often suffer from many of the same bureaucratic rigidities as public ones. Likewise although market-based interaction between providers and patients is most common in the private sector, short-term market exchanges in the form of user fees are pervasive in the public sector in many low income countries".

[114] The public sector is the main source of health funding in all OECD countries, except the United States, Mexico and Korea (OECD, 2003). The public share of health spending stood at 72,2% on average across OECD countries in 2000. It accounted for more than 80% of total health expenditure in several countries, including the Czech Republic, Denmark and the United Kingdom. The remaining 20% where spread between private insurance and out-of-pocket payments. Furthermore, there is evidence suggesting that there has been a convergence in the share of public spending over the past three decades (OECD, 2003). Many countries which started with a relatively high public share in 1970 had a lower public share in 2000 (*e.g.*, the Czech Republic, Norway and the United Kingdom), while several countries which started with a low public share in 1970 have seen this share increase over time (e.g., the United States, Greece and Portugal).

advantage of lacking a profit-maximising motif, and enabling greater political control over health service delivery, ensuring that transactions respect given priorities among interventions and patients. However, if the public sector holds the monopoly of health care provision it will lack competitive incentives for efficiency and innovation, which will tend to be aggravated by the typical rigidities of public hierarchical organisations reflected in the vested interests of bureaucrats and limited adaptability capacity to changing priorities and strategic orientations. This in turn leads to our second hypothesis: modes of governance that combine public and private provision and that integrate market-type mechanisms[115] into the operation of the public sector are expected to counteract the perverse incentives of pure profit maximisation, the monopolistic and bureaucratic failures of exclusive public provision, and to maximise efficiency.

Recent reforms in some OECD health systems embody some of the underlying prescriptions of this second hypothesis (Docteur and Oxley, 2003). One of the key areas of reform has consisted in the separation of purchasers and providers within public integrated systems and, more generally, the strengthening of purchasers' agency role within the health-care system. In parallel, there has been a shift towards more independent producers with greater management independence and

[115] Similar transformations in the way public health providers operate to those discussed for publicly managed schools in the previous chapter. These include changes in the incentive environment of organisations as well as modifications to the distribution of decision-making control, revenue rights, and financial risk among the different players. WHO (2000, p.63) highlights some of the key issues at stake:

• The degree of autonomy (decision rights) that the organisation has vis-à-vis its owners, policy-based purchasers such as insurance funds, the government, and consumers. Critical decision rights include control over input mix and level, outputs and scope of activities, financial management, clinical and non-clinical administration, strategic management, and market strategy (where appropriate).

• The degree of accountability. As decision rights are delegated to the organisation, the ability of governments to assert direct accountability (through the hierarchy) is diminished. When autonomy increases, accountability must be secured by shifting from hierarchical supervision to reliance on monitoring, regulations, and the economic incentives embedded in contracts.

• The degree of market exposure or revenues that are earned in a competitive way rather than through a direct budget allocation. Market participation need not imply out-of-pocket financing; it is preferable for provider organisations to compete for prepaid revenues. When governments bail out organisations that run deficits or are indebted as a result of weak technical performance, they undermine the impact of market exposure.

• The degree of financial responsibility for losses and rights to profit (retained earnings and the proceeds from the sale of capital). This determines the financial incentive for managers and staff to economize. Under increased autonomy they, rather than the public purse, become the "residual claimant" on revenue flows, but such claims must be clearly spelled out and regulated.

• The degree of unfunded mandates. Where the share of total revenues earned through markets is significant, organisations are at financial risk because of the unrecoverable costs associated with requirements for which no funds are provided, such as care for the poor or very sick. Organisational reforms that increase autonomy should therefore be accompanied by complementary reforms in health financing to protect the poor.

responsibility. Contracting-out selected activities has also increased (the typical example is the Private Finance Initiative in the United Kingdom which allows the private sector to build hospitals and to operate all non-medical services within them under contract). A limited number of countries (the United States, the United Kingdom, Sweden, the Czech Republic and New Zealand) have experimented with greater competition among hospitals[116] as a means of inducing improvements in efficiency. However, in many cases competitive pressures and provider incentives were weak, purchasers lacked the skills and information to place enough pressure on providers for change, conditions of tight supply gave providers a strong market position, and policies were in place for too short a period to have a substantive impact. As a result, policies were often set back and, in some cases (New Zealand, for example) reforms were almost completely reversed. While attempts at active competition in health-care markets have been curtailed, some of the underlying elements of these reforms nonetheless remain. All countries appear to have maintained contracting arrangements, even if they have become longer-term in nature and place greater emphasis on co-operation than on head-to-head competition.

In order to test our two hypotheses regarding the connection between modes of governance and health efficiency, we draw on the typology of OECD health systems provided by Docteur and Oxley (2003, p.9 and 10)[117]:

***

Although there is considerable variation within systems, OECD countries can be classified as generally consistent with one of the three approaches described below. It is important to recognise that elements of more than one of these approaches exist in most countries (even if one form is dominant) and that the dominant model has tended to shift under the force of reforms:

The *public-integrated model* combines on-budget financing of health-care provision with hospital providers that are part of the government sector. These systems, which merge the insurance and provision functions, are organised and operated like any government department. Staff is generally paid on salary (although, in some cases, doctors can have private patients as well) and they are most often public-sector employees. Ambulatory doctors and other health-care professionals can be either

[116] Healthcare providers competing for the customers of health-care purchasers on the basis of price.
[117] Few attempts have been made to create an *empirical* categorisation of health care systems. This is mainly explained by the heterogeneity of existing systems, and the coexistence of various models within each system. The same cannot be said of the "industry" on *theoretical* typologies of health systems – for a recent survey see Burau and Blank, 2006.

public employees or private contractors to the health-care authority, with a range of remuneration packages.

In the *public-contract* model, public payers contract with private health-care providers. The payers can be either a state agency or social security funds. Single-payer arrangements have a stronger position *vis à vis* providers (as in the public integrated model) and tend to have lower administrative costs than do multiple payer systems. In many public-contract systems, the private hospitals and clinics are run on a non-profit basis. Independent private contractors generally supply ambulatory care. In the past, payment of providers has been often on an *ex post* basis for services provided, although contract arrangements have been evolving.

A *private insurance/provider model* uses private insurance combined with private (often for-profit) providers. Insurance can be mandatory (Switzerland) or voluntary (the United States), and in the case of the latter, affordable insurance may not be available to some individuals. Payment methods have traditionally been activity based…Under these arrangements, insurers selectively contract with competing providers and restrict patient choice of providers and services.

***

We did not embrace this categorisation in its pure form as it does not take into consideration the recent transformations of public integrated systems. In the same study, Docteur and Oxley (2003, p.10, p.32) provide additional information that enables us to catagorise health systems allowing for those transformations:

***

Broadly speaking, public-integrated systems exist in the Nordic countries, Australia (public hospitals), Italy, Greece and Portugal and, before reforms of the early 1990s, the United Kingdom. New Zealand introduced a purchaser-provider split in the 1990s similar to developments in the United Kingdom, but it has since moved closer to an integrated model following reforms in 2000. Canada, most of the remaining Continental European countries, Japan, and, now, the United Kingdom and, to some extent, New Zealand, belong to the public-contract category.

***

A first area of reform concerns the separation of purchasers and providers within public integrated systems and, more generally, the strengthening of purchasers' agency role within the health-care system....A significant number of countries with integrated systems have now moved in this direction (Australia, United Kingdom, New Zealand, Sweden, Italy, Portugal and, more recently, Greece). More active purchasing has also occurred in countries with public contract models (Germany, Belgium).

***

By complementing this information with OECD health care reviews and country profiles of the European Observatory on Health Systems and Policies, we compiled the following classification:

i) Public integrated models without active purchaser-provider split (PPS): Norway, Denmark, Finland, Spain, France, Iceland.

ii) Public integrated models with active purchaser-provider split: the United Kingdom, New Zealand, Australia, Italy, Greece, Portugal, Sweden.

iii) Public contract models: Austria, Luxembourg, Canada, Germany, Belgium, Japan, Netherlands, Ireland.

iv) Private insurance models: the United States, Switzerland.

According to the first hypothesis laid out in this section, models i and ii should outperform models iii and iv in terms of efficiency. According to the second hypothesis model ii should be the optimal mode of governing health services, and therefore should outperform all the other models. We investigate the empirical validity of these hypotheses in section 6.5. In the next section we discuss the methodological assumptions that underlie the efficiency estimates examined in section 6.5.

6.4 Health Efficiency: Methodological Issues

Measuring the efficiency of health services is an exercise fraught with difficulties. Problems arise in measuring the performance of health systems, in identifying the entire range of variables that influence the health status of individuals; in understanding the pathways whereby those variables affect health status and in particular the way they interact with each other in doing so; and in finding empirical data for all these variables. With such a litany of problems, any interpretation of results needs to be cautious.

Individual health depends on medical and non-medical sets of inputs. The first set includes inputs under control of the health system (part of its domain of accountability). Examples of such inputs include human resources, drugs, and physical infrastructure. Since detailed information on these items is available for only a tiny handful of countries, it is common to use financial variables (e.g. total expenditure per capita) as an aggregate proxy. The second group includes inputs that, by their own nature, escape the control of health systems, and for which therefore they

should not be held accountable. The list of non-medical determinants of health is long: GDP per capita, income inequalities, educational attainment, employment status, quality of housing, diet, smoking, alcohol consumption, physical activity and other lifestyle factors, genetic inheritance, etc. In theory the effects of these variables on health should be controlled for. In practice it is very difficult to take proper account of these affects. Many of these variables interact with each other[118], with medical inputs, and with individual health in ways that are still poorly understood. Furthermore, data on behavioural variables (e.g. diet, smoking, alcohol consumption) and genetic inheritance is not yet available at the international level. Not controlling for the effects of these variables on health status remains a serious limitation of most studies on health efficiency given the high risk of model misspecification (by omission of variables). The results presented in the next section do not circumvent this risk.

WHO (2000) defined three core policy goals for assessing the performance of health systems: (i) average level of population health and health inequalities in the population; (ii) average level of responsiveness of the health system and inequalities in responsiveness within the population (quality of interaction between consumers and providers with particular emphasis on such issues as dignity, autonomy, confidentiality, promptness of attention, access to social support networks, basic amenities, and choice of provider); and (iii) fairness in the way the costs of the health system are distributed across households (situations where households are forced to pay a catastrophic share of their non-subsistence income to the health system; cases where households in similar circumstances contribute very different shares of their non-subsistence income to the health system; and finally, the extent to which the poor contribute a larger share of their disposable income for health in comparison to the rich).

The levels of health and responsiveness define the quality of the health system whereas the distributions of health, responsiveness, and financial burden define the degree of equity embodied in the system. WHO (2000) drew on an internet survey of informed respondents for determining the weights to be attached to these five outcomes in the construction of a composite indicator of health system

[118] Education attainment, for example, is often correlated with GDP per capita, which in turn is correlated with health expenditure per capita (OECD, 2002).

performance[119]. However, this indicator has been extensively criticised due to poor methodological procedures in the collection and treatment of data (see for example Williams, 2001). This criticism has had a strong echo in the literature on health sector efficiency, which generally relinquishes the use of WHO's composite index as a measure of health system's outcome in favour of one-dimensional indicators such as life expectancy, infant and maternal mortality or Disability-Adjusted Life Expectancy[120].

Difficulties in defining, measuring and modelling the inputs and outcomes of health systems make the measurement of efficiency a delicate empirical issue. Not surprisingly the existing literature on the topic is rather limited, particularly when it comes to international comparative studies[121].

Some of the most interesting contributions in the field include the work of Gupta and Verhoeven (2001) who applied FDH to a panel of data from 85 countries for various years between 1984 and 1995. The frontier was defined for life expectancy and infant mortality, while average per capita government spending (the private sector was ignored) on health and education were used as inputs to the production process. The authors controlled for the initial level of development by integrating the initial GDP per capita in the analysis. Evans, Tandon, Murray and Lauer (2000) estimated an efficiency frontier for 191 countries using an extension of least squares to panel data from 1993 to 1997. DALE, the chosen output indicator, was regressed on total health expenditure per capita, and a measure of educational attainment in the adult population as a control variable for non-health-system direct inputs. While studying the efficiency of the Portuguese health system St. Aubyn (2002) applied FDH and corrected least squares to OECD countries. Two different output indicators (DALE

[119] The average weights from over 1 600 responses were 24% health, 25% health inequality, 13% responsiveness, 16% responsiveness inequality, and 22% fairness in financial contribution. These results are consistent with those collected by Gakigou *et al.* (2003) who made the first attempt to measure preferences of the general public on the relative importance of the goals of health systems drawing on data from 51 countries and more than 53 000 respondents.

[120] DALE is estimated from three kinds of information (Murray *et al.*, 1999): the fraction of the population surviving to each age, calculated from birth and death rates; the prevalence of each type of disability at each age; and the weight assigned to each type of disability, which may or may not vary with age. Survival at each age is adjusted downward by the sum of all the disability effects, each of which is the product of a weight and the complement of a prevalence (the share of the population not suffering that disability). These adjusted survival shares are then divided by the initial population, before any mortality occurred, to give the average number of equivalent healthy life years that a newborn member of the population could expect to live.

[121] This contrasts with the study of technical efficiency of specific health care activities (e.g. secondary care and nursing home care; primary care delivery; and individual physician efficiency), where numerous contributions exist (for a comprehensive survey of these studies see Hollingsworth, 2003).

and infant mortality) were separately considered, and total health expenditure per head was selected as the appropriate input variable. In some specifications the percentage of the population without higher secondary or tertiary education entered as a control variable. Hollingsworth and Wildman (2002) computed cross efficiency measures (a means of validating Data Envelopment Analysis scores through different weighting schemes) using data from 30 OECD countries (again the selected variables were DALE, total health care expenditure per capita, and average years of schooling in the adult population). Afonso, Schuknecht and Tanzi (2005) incorporated the health system in their international comparison of public sector efficiency. They used FDH for comparing the relative efficiency of the public sectors of 23 industrialized OECD countries in 2000. Public spending as a percentage of GDP in 2000 was taken as the relevant input variable. A composite index of opportunity indicators (public administration, education, health and public infrastructure) and standard musgravian indicators (distribution, stability and economic performance) measured overall public sector performance. Within this composite index, health sector performance was appraised through an aggregate measure of infant mortality and life expectancy. Finally, Afonso and St Aubyn (2004) addressed the efficiency of health services in OECD through a combination of FDH and DEA with financial and physical measures of inputs.

In estimating the efficiency of health systems we applied DEA and SFA to four different input-output datasets for a sample of 23 OECD countries[122]. The results of such exercise are used in section 6.5 to test the effect of modes of governance on efficiency.

In the different specifications of the efficiency frontier we did *not* control for non-health system inputs for four different reasons. First, following the position of the World Health Organisation (see for example Evans, D., Tandon, A., Murray, C. and Lauer, J., 2003) health systems should be held accountable for some of these factors (e.g. discouraging tobacco consumption), which implies that they should not be included as separate inputs. Second, as we mentioned earlier variables such as GDP per capita or adult education are highly collinear with health expenditure per capita (one of the input variables used), bringing an insignificant contribution to the

[122] The Czech Republic, Korea, Poland, Mexico, Turkey, Hungary, and Slovak Republic were left out of the analysis because they were deemed to create a bias in the efficiency rankings as a result of both low expenditure levels and results.

estimation of the efficiency frontier. Third, there is a lack of reliable international data for items such as dietary habits, smoking and alcohol consumption. Fourth, limiting our analysis to OECD countries confers a basic degree of homogeneity to the sample that mitigates the potential pitfalls (omission variable bias) of not controlling for some of these non-health system inputs.

In the first dataset, infant survival rate (ISR)[123] and life expectancy are the two output variables selected (both in 1999) and per capita health expenditure in purchasing power parities (PPPs) is taken as the input variable (values for 1998)[124]. Because SFA in its primal form is not compatible with more than one output indicator, we aggregated the ISR and life expectancy indicators into one composite index. Equal weights (50/50) were accorded to the two components[125].

In the second dataset, practising physicians (density per 1000 population), practicing nurses (density per 1000 population), and total in patient care beds per 1000 population are the three input variables used. ISR and life expectancy are maintained as the two output indicators[126]. Important physical inputs such as drugs, equipment and technology are left out of the analysis due to unavailability of data. This constitutes an important limitation to this particular specification of the efficiency frontier and stresses the importance of testing the effects of modes of governance on efficiency through a range of different input-output datasets and methodological approaches to the estimation of efficiency.

In the third dataset, disability-adjusted life expectancy (DALE) is the output variable and the usual total health expenditure per capita in PPPs the input variable. Both variables refer to 2002[127].

In the fourth dataset, WHO's (2000) composite index discussed above is the selected output variable and total health expenditure per capita in PPPs the input variable (both input and ouput variables concern 1997[128]).

---

[123] (1000-Infant Mortality Rate)/Infant Mortality Rate, i.e. the ratio of children that survived the first year to the number of children that died.

[124] Data were extracted from OECD (2002). This same dataset is also used by Afonso and Aubyn (2004).

[125] Life expanctancy and infant mortality are two of the most commonly used indicators of health status. In the absence of any theoretical weighting criteria we decided to combine both indicators using equal weights.

[126] Since only physical inputs are being used, the ensuing estimates concern only technical efficiency. Data were extracted from OECD (2002). This dataset is also used by Afonso and Aubyn (2004).

[127] Data on DALE were extracted from WHO (2003), and health expenditure was obtained from OECD (2004c).

[128] Data extracted from WHO (2003).

6.5 Results

The ensuing efficiency rankings displayed considerable sensitivity to the employment of different methodologies in the estimation of the efficiency frontier (DEA and SFA), to the use of different input-output datasets, as well as to the input/output orientation of the analysis. Given the extensive number of methodology/dataset/input-output-orientation combinations and the lack of a clear efficiency/inefficiency pattern amongst the results, an aggregate league table is presented in this section (Table 6.1)[129].

A simple arithmetic average of input and output efficiency scores was calculated. In order to facilitate the compilation we normalized the scores of each ranking and set the individual averages to one. However, it should be noticed that this table is meant to be merely illustrative of the average relative performance of each country across the various methodologies/datasets. No other meaning or interpretation should be derived from these figures as they mix together efficiency estimates based on different variables and methodological assumptions.

Table 6.1: Average input and average output efficiency

Country	Avg input efficiency	Country	Avg output efficiency
Japan	1.34	Japan	1.03
Spain	1.29	Spain	1.03
Portugal	1.29	Sweden	1.02
Greece	1.28	Greece	1.01
United Kingdom	1.21	Iceland	1.01
New Zealand	1.11	United Kingdom	1.01
Sweden	1.11	France	1.01
Finland	1.10	Italy	1.01
Ireland	1.04	Finland	1.00
Iceland	1.01	Norway	1.00
Italy	0.99	Portugal	1.00
Australia	0.96	Canada	1.00
Canada	0.94	Switzerland	1.00

[129] Hollingsworth and Wildman's (2003) results were also integrated in this aggregate ranking. The authors estimated health service efficiency of OECD countries using cross efficiency measures designed to validate data envelopment scores through different weighting schemes for inputs and outputs. In the estimation process, the output (measured in DALEs) was considered a function of the physical inputs into the health system (proxied by health expenditure per capita), and non-health inputs (measured by the average years of schooling in the adult population). The underlying data were extracted from WHO (2000).

Belgium	0.92	Austria	1.00
Austria	0.92	Australia	1.00
France	0.90	New Zealand	1.00
Norway	0.90	Belgium	1.00
Denmark	0.88	Luxembourg	1.00
Luxembourg	0.84	Germany	0.99
Netherlands	0.83	Ireland	0.99
Germany	0.75	Netherlands	0.99
Switzerland	0.73	Denmark	0.99
United States	0.61	United States	0.95

Notwithstanding the interpretative limitations of table 1, it is still interesting to note that Japan, Spain, Greece, the United Kingdom, Sweden, Finland and Iceland come out of this aggregative exercise as examples of good performance, irrespective of whether the analysis is input/output orientated. Conversely, Denmark, Luxembourg, the Netherlands, Germany, Switzerland, and the United Sates show evidence of poor performance in the provision of health care services, both in terms of input and output efficiency.

In order to assess the hypothesis laid out in section 6.3, we regressed the individual DEA and SFA efficiency scores of our 4 input-output datasets on three dummy variables that capture the four modes of governance discussed in that same section. Hollingsworth and Wildman's (2003) DEA cross efficiency scores were also integrated in this exercise. The results are presented in Table 6.2.

Table 6.2: Multiple regression analysis, modes of governance (x) vs efficiency (Y)

Methodology/Dataset	**Input Efficiency**		
	Public Contract Models	Public Integrated Models *with* PPS	Public Integrated Models *without* PPS
DEA expenditure pc vs life expectancy and ISR	0.285**	0.464***	0.400***
DEA doctor, nurses and beds vs life expectancy and ISR	-0.224	-0.082	-0.154
DEA expenditure pc vs disability life expectancy (2002)	0.264**	0.398***	0.294**
DEA expenditure pc vs WHO 2000 composite indicator	0.277**	0.394***	0.273**
DEA cross efficiency scores (Hollingsworth *et al.*, 2003)			
All efficiency tables combined (table 1)	0.277**	0.466***	0.345**
Methodology/Dataset	**Output Efficiency**		
	Public Contract Models	Public Integrated Models *with* PPS	Public Integrated Models *without* PPS
DEA expenditure pc vs life expectancy and ISR	0.001	0.011	0.013
SFA expenditure pc vs combined ISR and life expectancy	0.087***	0.090***	0.106***

DEA doctor, nurses and beds vs life expectancy and ISR	-0.020	-0.006	-0.020
SFA doctor, nurses and beds vs combined life expectancy and ISR	0.046	0.059	0.076**
DEA expenditure pc vs disability life expectancy (2002)	0.007	0.012	0.014
SFA expenditure pc vs disability life expectancy (2002)	0.021**	0.026***	0.023**
DEA expenditure pc vs WHO 2000 composite indicator	0.002	0.003	0.001
SFA expenditure pc vs WHO 2000 composite indicator	0.026***	0.029***	0.027***
DEA cross efficiency scores (Hollingsworth *et al.*, 2003)	0.056**	0.073***	0.061**
All efficiency tables combined (table 1)	0.025**	0.033**	0.034**

Notes:
The private insurance/ private provider model was taken as the benchmark in each regression.
*** means the estimate is significant at the 1% level; ** means the estimate is significant at the 5%level; * means the estimate is significant at the 10% level.

We obtained significant evidence in two thirds of the regressions analysed. Overall, the public integrated system *with* purchaser-provider split (PPS) is the model that generates higher average efficiency scores in regressions run over 7 different methodologies/datasets, whereas the public integrated systems *without* active PPS does better in the remaining 3 significant regressions.

Regardless of whether the analysis is input or output orientated, the estimates of the dummy variables in the significant regressions are positive with no exception, meaning that both public contract models, as well as public integrated models with and without active PPS systematically outperform private insurance models (our selected benckmark). Public contract models are linked with an average input efficiency score 26.41 to 28.54 percentage points higher than that of private insurance models. For public integrated models with and without PPS that interval is of 39.38 – 46.607 and 27.34 – 39.95 respectively. When efficiency is measured in an output-oriented way, public contract models are estimated to achieve average efficiency scores 2.13 – 8.67 percentage points above private insurance models, while for public integrated systems with and without PPS these figures vary between 2.57 – 9 and 2.3 – 10.62 percentage points respectively.

When the effects of the four modes of governance are tested upon the average input efficiency scores calculated earlier on (table 2), public integrated models with PPS emerge as the top performers with an average score 46.61 percentage points above that of private insurance models, followed by public integrated systems without PPS (34.46) and public contract models (27.7). As for output efficiency, public integrated models without PPS occupy the first position 33.9 percentage points above private insurance models, followed by public integrated system with PPS (33) and public contract models (24.8).

6.6 Conclusion

This chapter investigated the research hypothesis formulated in the seventh section of chapter 3 within the context of health care services. According to this hypothesis, the organisational structure that maiximises efficiency is the one that confines government interference to the specific attributes of transactions that markets fail to address, and in a way that allows for risks of government failure. Applying this general hypothesis to the health sector led us to formulate two context-specific hypotheses.

First, problems of moral hazard and adverse selection hinder the efficiency of free competitive health insurance markets, and prompt extensive government intervention in the funding of health care. Hence our first research hypothesis was that modes of governance based on direct, short-term market interactions between patients and providers, or relying mainly on unregulated private insurance were expected to be relatively inefficient.

Second, the combination of profit-maximisation with imperfect information propagates perverse incentives that lead competitive markets to failure. Public sector monopoly of health care provision, on the other hand, lacks competitive incentives for efficiency and innovation, and suffers from the typical rigidities of public hierarchical organisations (risk of government failure). This in turn led us to our second hypothesis: modes of governance that combine public and private provision and that integrate market-type mechanisms into the operation of the public sector are expected to counteract the perverse incentives of pure profit maximisation, the monopolistic and bureaucratic failures of exclusive public provision, and to maximise efficiency.

The results presented in section 6.5 corroborated the two hypothesis. Public integrated systems with active purchaser provider split came out of our analysis as the most efficient mode of organising health services. While the introduction of market-type mechanisms into public integrated models seemed to increased efficiency, further movement towards a market-inspired mode of governance appeared to have detrimental effects on efficiency (public contract systems and private insurance/private provider systems were systematically outperformed by public integrated systems).

Chapter 7: Modes of Governance and Social Protection Efficiency

7.1 Introduction

This chapter examines our research hypothesis in the context of social protection. The governance of social protection systems takes us through an analytical route significantly different from the ones pursued in the two previous chapters for two different reasons. First, the bulk of social protection activity takes the form of social transfers in cash to individuals in (or at the risk of) poverty, or in pre-defined states of nature (families with children, etc). Therefore, contrary to education and healthcare, in this case there is no service being effectively provided other than an insurance service. The governance of social protection is thus confined to the funding and (re)distributional elements of an organisational arrangement and does not include elements of service provision in a technical sense. Second, private social insurance markets face a litany of problems they undermine their viability as a mode of governance for social protection. Again contrary to education and healthcare, here the relevant choices are not to do with the mix of public/private involvement in social protection, but with the coverage and nature of public social protection systems.

The efficiency of social protection systems is an essentially neglected issue in the academic literature. Most of the existing studies have focused on assessing the impact of social security and taxation policies on income distribution without considering simultaneously the financial effort involved in these policies (Esping-Andersen, 1990; Myles, 1989; Palme, 1990; Bradshaw et. al, 1993; Atkinson, Smeeding and Rainwater, 1995; Forster, 1993; Kenworthy, 1998). When reference to efficiency is made it tends to be confined to the definition of a ratio between a measure of the impact of taxes and transfers on income distribution (or specifically on poverty alleviation) and a measure of social protection effort put in by the state, which technically corresponds to the notion of productivity rather than efficiency (e.g. Beckerman, 1979; Beckerman and Clark, 1982; Mitchell, 1991; Mitchell, Harding and Gruen, 1994).

The purpose of this chapter is to develop an analytical framework to estimate formally the efficiency of social protection systems and to investigate its causes. To do so we estimate a stochastic efficiency frontier for public non-pension social transfers in 19 OECD countries using panel data from the mid-1990s and the year 2000. The targeting of transfers towards the bottom deciles of the working age

population is thought to be the key explanatory variable of efficiency and its effect is estimated together with efficiency itself in a single-stage procedure.

In addition to measuring and explaining efficiency, we use our results to test a common underlying claim of two different strands of literature[130] which suggests that there is a trade-off between the size of transfers and their targeting. We find no empirical support for such a view.

After measuring efficiency and tracing it back to the targeting of transfers, we investigate the governance arrangements that lead different social protection systems to achieve different targeting patterns. Drawing on Kraus' (2004) organisational typology of modern social security systems, we find the combinations Beveridge-con-social-assistance and Bismarck-without-social-assistance to be associated with, respectively, the highest and lowest levels of efficiency.

The chapter proceeds as follows. Section two presents a brief overview of the contextual conditions that have shaped the evolution of social protection systems, as well as the attributes of social protection that call for government intervention. Section three discusses the methodological issues that underpin the measurement of social protection efficiency. Section four provides the main efficiency results and investigates the alleged link between the size of transfers and their targeting. Section five examines the effects of different social security governance arrangements on efficiency. Section six provides some concluding remarks.

7.2 The Economics of Social Protection

In most advanced economies, social protection systems emerged and developed in a context characterised by high economic growth, low unemployment, strong national states, and limited international competition faced by domestic firms. As conditions evolved though, these systems generated high levels of expenditure, which weighed heavily on public budgets. Their roots can be found in decisions made especially in the 1960s and 1970s to create and extend public programmes in the fields of education, health, old-age and other areas. Once established, these programmes created spending obligations that have proved hard to control. As economic growth decelerated and unemployment rose, the financial sustainability of

[130] Korpi, 1983; Korpi and Palme, 1998; and Mahler and Jesuit, 2004, on the one hand, and Tanzi and Schuknecht, 1997a, 2000, on the other.

social protection systems was called into question. Additional pressures to reform also came from perceptions that the design features of some of the social programmes had perverse effects on individuals' incentives to work and save, which could compromise economic growth. This section begins with a brief survey of the contextual conditions that have shaped the evolution of social protection since the last quarter of the eighteenth century (section 7.2.1).

Social protection systems define the collective intervention of society to shield citizens from a wide range of vulnerabilities, sustain their well-being and enhance their capabilities in managing risks. These systems have proved successful in ensuring that the structural adjustment associated with continued economic growth did not translate into unacceptable consequences for the well-being of vulnerable groups. Their contribution to maintaining a policy environment that facilitates sustainable economic growth makes them an *asset of society* (Ercole and Salvini, 2003), generating benefits that accrue to the populations at large and hence must be borne collectively. In section 7.2.2, we discuss the attributes of social protection whose efficient coordination requires that public central coordination supersede the market.

7.2.1 The Growth of Government and the Development of Social Protection

At the beginning of the twentieth century government spending in today's industrial nations accounted for less than one-tenth of national income. In the 1990s the government's share of output was roughly half. A combination of both reactive actions to unpredicted events (such as wars and depressions) as well as voluntary policies (in tune with the prevailing ideology on the role of the state) has been responsible for such increase in the economic weight of government.

Classical economists favoured a small allocative role for government. As a reflection of this limited role, in the 19th Century the state provided law and order and (mostly in the second half of the 19th century) became involved in primary education and in the provision of basic infrastructure. Governments were expected to balance their budget in peace times, and they had hardly any role in income distribution or stabilisation (Bastiat, 1944-45; Buchanan, 1985). By modern standards, public

spending was low as a share of gross domestic product (10 percent on average) and did not change much between 1870 and World War I[131] (Table 7.1).

Between 1913 and the beginning of World War II, public spending grew substantially. As a consequence of World War I, governments increased their ability to tax and consequently, to spend (Peacock and Wiseman, 1961). Also, in the late 19th century, economists like Schmoller and Wagner had started advocating a distributive role for the state. As a consequence of World War I and the Great Depression, governments of several countries started to develop extensive and largely redistributive spending programs (such as those associated with the New Deal in the United States). Social security systems with limited coverage and benefit levels were introduced in many countries between 1880 and 1930 and, over time, added to budgetary obligations of governments. Average public spending rose from about 12 percent in 1913 to about 18 percent after World War I, and to over 22% before World War II.

The post-war period, and in particular the 1960s and 1970s, witnessed a growth in public spending, unseen during peace times. This growth resulted from the enthusiasm about what government could do with respect to redistribution, stabilisation and even allocation. This was the period when 'wars on poverty' or the welfare state were launched. By 1980 average spending in industrial countries had increased to over 40 percent of GDP; by 1995 it had reached 46 percent of GDP[132].

Until World War II, most of the growth in government spending had been on goods and services. After World War II, and especially after 1960, however, subsidies and transfers, especially in cash, were the driving force behind government growth. In 1870, transfers and subsidies amounted to 1 percent of GDP or 10 percent of total spending. By 1937, this share had risen to over 4 percent of GDP, largely due to the introduction of basic social insurance and to unemployment benefits, the latter as a consequence of the Great Depression. Between 1960 and 1980, subsidies and

[131] "The prevailing attitude of economists vis-a-vis the role of government at the turn of the century can be illustrated by citing an influential French economist who wrote that the level of taxation could be called 'moderate' when the ratio of all taxes in national income was 5 to 6%. Taxes became 'heavy' when the ratio was 10 to 12%. Beyond that percentage the rate of taxation would be 'exorbitant' and would have serious consequences for the growth of the country and the liberty of its citizens and its industry." (Tanzi and Shuknecht, 1997, p. 398).

[132] "When the overall developments between the late nineteenth century and the late twentieth century are compared, it is noticed that half of the growth in government expenditure – from 10 percent of GDP around 1870 to 28 percent in 1960 – occurred during the two world wars. Expenditure growth to 46 percent of GDP in the thirty-six after 1960 equalled the expenditure growth in the previous ninety years even though the post-1960 period was free of major wars or depressions." (Tanzi, 2000, p. 20)

transfers increased from less than 10 percent to 21.2 percent of GDP, and in 1980 subsidies and transfers absorbed 50 percent of total public spending. After 1980, the expansionary trend continued but at a much slower pace. Expenditure on transfers and subsidies increased from an average of 21.4 percent of GDP in 1980 to 23.2 percent of GDP in 1995, slightly more than half of total government expenditure. While transfers continued to grow, producer subsidies, however, started coming down in a majority of countries.

Table 7.1: The growth of general government expenditure, 1870-1995 in percentage of GDP (adapted from Tanzi and Schuknecht, 1997b, p. 397)

	Later 19th Century about 1870	Pre-World War I 1913	Post World War I 1920	Pre-World War II 1937	Post World War II 1960	1980	1990	1995
General government for all years								
Australia	18.3	16.5	19.3	14.8	21.2	31.6	34.7	37.1
Austria	-	-	14.7	20.6	35.7	48.1	48.6	52.8
Canada	-	-	16.7	25.0	28.6	38.8	46.0	46.2
France	12.6	17.0	27.6	29.0	34.6	46.1	49.8	53.7
Germany	10.0	14.8	25.0	34.1	32.4	47.9	45.1	49.5
Ireland	-	-	18.8	25.5	28.0	48.9	41.2	42.0
Japan	8.8	8.3	14.8	25.4	17.5	32.0	31.7	35.6
New Zealand	-	-	24.6	25.3	26.9	38.1	41.3	34.7
Norway	5.9	9.3	16.0	11.8	29.9	37.5	53.8	49.2
Sweden	5.7	10.4	10.9	16.5	31.0	60.1	59.1	66.2
Switzerland	16.5	14.0	17.0	24.1	17.2	32.8	33.5	39.4
United Kingdom	9.4	12.7	26.2	30.0	32.2	43.0	39.9	43.4
United States	7.3	7.5	12.1	19.7	27.0	31.8	33.3	33.3
Average	10.5	12.3	18.7	23.2	27.9	41.3	42.9	44.9
Central government for 1870-1937, general government thereafter								
Belgium	-	13.8	22.1	21.8	30.3	58.6	54.8	54.9
Italy	11.9	11.1	22.5	24.5	30.1	41.9	53.2	51.9
Netherlands	9.1	9.0	13.5	19.0	33.7	55.2	54.0	50.9
Spain	-	11.0	8.3	13.2	18.8	32.2	42.0	44.3
Average	10.5	11.2	16.6	19.6	28.2	47.0	51.0	50.5
Total average	10.5	11.9	18.2	22.4	27.9	42.6	44.8	46.1

The rapid expansion of public expenditure between 1960 and 1980 is remarkable because it occurred when most countries were not engaged in war effort; there was no depression, and the demographic developments were generally fiscally friendly. It essentially reflected a change in attitude towards the role of the state. The 1960s and 1970s was the heyday of Keynesianism and the time when governments were perceived by many to be efficient in allocating and redistributing resources and in stabilising the economy. This was also the period when basic social security

systems acquired some of the characteristics of the welfare state. In many countries the social security systems expanded their activities outside the field of insurance for old age, and the link between benefits and contributions became more tenuous for individual contributors. For the most part, this growth was not induced by technical factors, such as declining government productivity or ageing populations, but by political decisions that extended services, thus turning limited social safety nets into universal social benefits.

The current policy commitments imply a considerable upward dynamic in some of the social expenditure categories, and could become an important threat for future budgetary stability. In future years "technical factors" such as ageing of the population will exert considerable pressure on public spending if current policies stay unchanged. This is one of the main reasons why current policies are being exposed to increasing pressures for change and efficiency considerations have become so popular in this field.

7.2.2 Social Protection and Public Central Coordination

In a world of perfect knowledge and perfect foresight, with well functioning capital and insurance markets, there would be no reason for the state to play an extensive role in social protection. Voluntary decisions, insurance contracts, and the saving instruments created by markets would suffice. Under such circumstances, government intervention would only be required at a residual level, as a means of last resort, providing a social safety net to those whose total income over the entire life-cycle is insufficient to sustain an adequate living standard[133]. The assurance of a minimum living standard for part of the population who would otherwise fall below the poverty line, apart from being justifiable on the basis of social conventions, such as moral criteria, or social solidarity, is also justifiable in terms of economic efficiency, given the externalities usually involved[134] (Spicker, 2000).

In the real world, however, insurance markets for social protection are known to be prone to moral hazard and adverse selection problems that lead them to

[133] In this case, income is redistributed in a way that is not actuarially linked to the contributions made to the system by the recipient.

[134] Social protection can be regarded as a "good" whose consumption generates positive effects that spill over to society as a whole. Some of these effects result from the interaction with education, health, and employment, which are known to influence not only the ones directly concerned but also to the rest of society where they are inserted.

inefficiency. Once insured, it is often difficult (if not impossible) for the insurer to monitor the behaviour of the individual who may not follow a proper conduct with regard to the particular risk subject to contract. If insured against unemployment, for example, a worker may find it advantageous to deliberately remain in that state, depending on the prevailing replacement rate. Furthermore, and as in the health sector, buyers of insurance generally have a clearer idea of the risk they represent than the insurer, so if the insurance premium is based on average risk, only individuals of average risk and above will be willing to buy the policy, which will bring financial losses to the insurer. On the other hand, if insurers somehow have knowledge on individual risk type, the market would be characterised by separating equilibrium, with full insurance for the good risks, but only partial or no insurance to the worse risks. Government intervention is then required so that a pooling equilibrium is reached by making the insurance compulsorily and regulating the ability of firms to accept/reject contracts.

Apart from moral hazard and adverse selection problems, there are other serious obstacles to the efficiency of insurance markets for social risks. One such obstacle is the uncertainty about the size of future claims and the resulting difficulty/impossibility in setting a realistic price for the insurance[135]. Another obstacle is risk correlation. Most of the risks social security systems attempt to tackle are correlated in time, which implies that the aggregate risk cannot be solved by pooling the individual risks. Public insurance schemes are not able to solve this problem either, but the state may have an edge in using its budget and public debt to shift resources across time.

More generally, it is often argued that, individuals, left to their own decisional criteria, tend to be sub-optimally provident, in that their saving pattern seems to partially ignore their future income needs. By short-sightedness or by deliberate strategic behaviour people may not save enough and become a burden to society, even if they are, in life cycle terms, above a minimum poverty level. It is thus only logical that if people benefit from some form of public insurance they should be coerced to pay for it. By so doing, the government is in effect compulsorily raising saving levels and smoothing consumption across the life cycle.

[135] This uncertainty is shaped by macroeconomic instability and by inflation in particular.

Efficiency, therefore, rests at the heart of the role fulfilled by public social protection systems (Lucena and Macedo, 1996). Given the failures of free insurance markets, public social security systems assist individuals with the process of smoothing income across the life cycle (such as in the case of old age pensions) or across states of nature (such as in the case of unemployment benefits). In both instances, the system can be set in a way that is actuarially fair for every one. Therefore, it is justifiable exclusively on account of efficiency considerations, irrespectively of any *a priori* purpose of redistribution. Pure income redistribution arises only when the state helps some of its citizens sustain a minimum living standard, without there having been a significant contribution on their behalf entitling them to such level of assistance, i.e. when individuals are not being given benefits strictly on the grounds of an insurance logic, but are in fact receiving them on the basis of social charity. Important a category this may be (and indeed it stands on itself as one fundamental pillar of the notion of welfare state), the fact is that it does not express the majority of operations carried out by social security systems in modern societies. It plays a residual role when compared with the benefits (in kind and in cash) received by individuals as part of an implicit insurance contract with the state. Perhaps more importantly, the "pure" income redistribution involved emerges as a by-product of the original goal of not allowing people to fall below a given living threshold, and not as a goal in itself.

7.3 The Efficiency of Social Protection Systems: Methodological Issues

As proposed by Farrell (1957, p. 254), "when one talks about the efficiency of a firm one usually means its success in producing as large as possible an output from a given set of inputs". In general, a decision-making unit (DMU) is said to be technically efficient if it generates the maximum possible output from a given set of inputs, or seen from a different angle, if it minimises the amount of inputs used to generate a given level of output.

Measuring efficiency usually involves three steps. First, inputs and outputs need to be properly defined and measured. Secondly, the set of feasible input-output combinations needs to be estimated (i.e. the production/efficiency frontier). Finally, comparing *actual* input-output combinations with *feasible* input-outputs yields efficiency estimates.

The relevant input/output variables for measuring the efficiency of social protection systems depend on the precise analytical definition of social protection one adopts. A broad conception of social protection usually includes a wide range of activities such as health, education, welfare services, and government transfer payments. Given the specificities that characterise both health and education services, often welfare services and governments transfer payments are studied separately as the specific domains of a narrower definition of social protection. In the following analysis we concentrate on this second restrictive conception of social protection and specifically on the role performed by social cash transfers[136].

Social protection systems define the collective intervention of society to protect citizens against vulnerabilities that may negatively affect their living standard, and enhance their capabilities in managing risks (Ércole and Salvini, 2003). This intervention encompasses two different components - collective insurance services and pure vertical redistributive assistance to the destitute (Lucena and Macedo, 1996).

The insurance component is designed to assist individuals with the process of smoothing income across the life cycle (e.g. old age pensions) or across states of nature (e.g. unemployment benefits)[137]. The practical effect of social transfers in both instances is to either prevent poverty and/or sustain living standards (a), or to mitigate/eliminate poverty (b).

The purely redistributive component helps citizens sustain a minimum living standard, without there having been sufficient contribution on their behalf entitling them to such level of assistance, i.e. when individuals are not being given benefits strictly on the grounds of an insurance logic, but are in fact receiving them on the basis of social solidarity. The practical reflection of social transfers in this case is essentially poverty mitigation/elimination (c).

Since poverty alleviation is the practical effect of transfers of type (c), (b) and to a significant extent (a), we will take it as the relevant output of social protection systems. In particular, our output variable is defined as the relative reduction of the

[136] Social protection provided "in-kind" represents a relatively small part of total social expenditure and is ignored in this study.
[137] In both instances, the system can be set in a way that is actuarially fair for every one. Therefore, it is justifiable exclusively on account of efficiency considerations, irrespectively of any *a priori* purpose of redistribution.

income poverty rate (IPR) before and after taxes and transfers for the working age population[138]:

$$\frac{IPR_{beforeTT} - IPR_{afterTT}}{IPR_{beforeTT}}$$

Equation 7.1

This is the most simple and frequently cited indicator of the state's redistributive impact on poverty (Mitchell, 1991), comparing the poverty rate associated with the actual distribution of income (disposable income) with the one that would result from the market, before any redistributive activity through the state's tax and transfer systems (market income).

The IPR is defined as the number of people in households below 50 per cent of the median disposable income in percent of the total population, after adjusting for household size. Incomes are recorded on an annual basis and all possible types of cash income are grouped into four categories (Förster and Pearson, 2002, p. 10):

i) Gross earnings: the salary income of the household from dependent employment (excluding employers' contributions to social security);
ii) Gross capital and self-employment incomes: financial gains, real estate rents, occupational pensions and all kinds of private transfers as well as self-employment incomes;
iii) Social security transfers: all kinds of cash transfers from public sources;
iv) Taxes: direct income taxes and employee social security contributions paid by households.

The IPR before taxes and transfers is defined in relation to market income (i + ii). The IPR after taxes and transfers is defined in relation to household disposable income (i + ii + iii - iv).

Having defined poverty alleviation as the relevant output of our narrow conception of social protection, it seems obvious to take the size of social transfers as the appropriate input variable for assessing efficiency. We extracted data on public social transfers as a percentage of GDP from the OECD Social Expenditure Database

[138] The data were kindly provided by Michael Forster and Marco Mira D'Ércole (OECD). A graphical representation of the data can be found in Forster and d'Ercole (2005, p.29).

(SOCX)[139]. Only cash benefits were considered since social protection provided *in kind* has no bearing on our output variable and therefore should be kept out of the analysis. The analysis is confined to the working age population and so old age pensions were not considered either[140].

The impact of redistributive policies on poverty alleviation depends not only on the size of social transfers, but also on the size of taxes and the progressivity of the structure of taxes and transfers (Barr, 1992). In particular, the extent to which a given amount of social transfers maximises its poverty reduction effect (i.e. the efficiency of transfers as we are measuring it) is explained by the share of social transfers effectively channelled towards the poor, after allowing for the counteracting effect of direct taxation levied upon them. We designate this share as the *net* targeting of transfers towards the poor (i.e. the gross share of transfers received by the poor - *gross* targeting - adjusted for the effect of direct taxes paid by the poor[141]):

$$\frac{Transf_{poor} - Taxes_{poor}}{Transf} \equiv \frac{Transf_{poor}}{Transf} * \left(1 - \frac{Taxes_{poor}}{Transf_{poor}}\right)$$

Equation 7.2

In estimating efficiency in section 7.4 we will take the share of social transfers received by the poorest 30 per cent of the population as the relevant indicator of the degree of targeting of transfers and include it in our model as an explanatory variable of efficiency[142].

---

[139] This mainly includes income support to the working-age population at risk from illness or loss of earnings (disability cash benefits, occupational injury and disease, sickness benefits, family cash benefits, unemployment benefits, housing benefits and other contingencies).

[140] In their essence, and if one excludes the relative minority represented by the non-contributory regime, old age pensions are direct compensations paid by the state in rendering an insurance service. Pensions are to a larger or lesser extent linked to the contributions of the beneficiaries, even in the so-called defined-benefit regimes. In the provision of this service, the state assists individuals smooth income across their life cycle. The prime function here is not to protect against states of nature where individuals might not be able to sustain an acceptable standard of life on their own earnings, but merely to enhance individuals' ability to save during their working years and transfer consumption to their retirement years. Since the underlying function of old age pensions is essentially distinct from that of most other social programs, we concentrate our analysis on transfers targeted at the working age population.

[141] The extent to which transfers are offset by taxation can be seen as one particular measure of fiscal churning (Palda, 1997).

[142] Data on the distribution of transfers is an adapted version of table A.5. in Forster and D'Ercole (2005) and were kindly provided by Michael Forster and Marco D'Ercole. The selection of the 30 percent threshold is justified on the grounds of it yielding data comparatively more reliable than that

Because countries depart from different market income poverty levels they will have different poverty reduction opportunities to explore and will exhibit different sensitivities to income redistribution policies. The adoption of similar redistributive policies[143] in countries with considerably different market income poverty rates will naturally result in dissimilar measures of poverty reduction. In order to allow for these dissimilarities we control for the initial poverty level by incorporating the reciprocal of the market income poverty rate as an input variable alongside social transfers.[144]

Data for all variables were collected for 1995 and 2000 and are presented in the statistical annex. Descriptive statistics of the variables are presented in Table 7.2.

Table 7.2: Descriptive Statistics

Variable	Mean	Std. Dev.	Min.	Max.
Poverty reduction	0.559	0.176	0.155	0.828
Social expenditure	0.067	0.025	0.023	0.128
Market poverty	0.188	0.035	0.115	0.266
Gross Targeting	0.473	0.115	0.258	0.785
Net Targeting	0.222	0.156	-0.210	0.545

With an appropriate set of inputs and outputs at hand, the next step is to determine the efficient set of input-output combinations against which individual social protection systems can be compared. Given that this frontier is not observable, one of two approaches can be followed. The first involves identifying the specific properties of each country's distribution of income (both before and after taxes and transfers); defining the set of feasible redistributive policies; and determining, for the observed level of social expenditure, the maximum poverty reduction effect achievable. Although ideal from a theoretical point of view, in practice this approach is extremely complex and demanding in terms of the information required, which currently renders it unfeasible for an international comparative study.

The second approach (the one we adopt here) is to estimate the frontier econometrically from a sample of observed inputs and outputs for different countries,

for the two bottom deciles of the population, whereas using the bottom four deciles would capture a significant part of the middle income groups.

[143] Here in the sense of similar levels of social expenditure (same level of input), roughly equally targeted (same level of our explanatory variable of efficiency).

[144] Under efficient conditions, a given level of social expenditure is expected to have a lower poverty reduction effect when the poverty rate before taxes and transfers is higher. This is meant to be taken into account through the inclusion of the control variable.

in different moments in time. The underlying procedure in this case is to compare a country with an ideal comparator constructed from information on other countries conducting similar operations. The purpose of efficiency analysis based on frontiers of this kind is to estimate a benchmark on the basis of real or observed performance, and not the theoretical maximum derived from a country-specific abstract production function[145].

As in the education chapter, we draw on the Battese and Coelli (1995) model for a stochastic frontier production function which is equivalent to the Kumbhakar, Ghosh and McGukin (1991) specification, with the exception that allocative efficiency is imposed, the first-order profit maximising conditions removed, and panel data is permitted. The model may be expressed as (Coelli, 1996b):

$$Y_{it} = X_{it}\beta + (V_{it} - U_{it})$$
$$i = 1,...,\mathrm{N}; t = 1,...,\mathrm{T}$$

Equation 7.3

Yi is the production (or the logarithm of the production) of the i-th firm;

Xi is a k×1 vector of (transformations of the) input quantities of the i-th firm;

β is an vector of unknown parameters;

t denotes the time period;

Vit are random variables assumed to be iid. N(0,σV2), and independent of Uit;

Uit which are non-negative random variables which are assumed to account for technical inefficiency in production and are assumed to be independently distributed as truncations at zero of the N(mit,σU2) distribution, where:

$$m_{it} = Z_{it}\delta$$

Equation 7.4

[145] Gouyette and Pestieau (1999) come close to a formal measure of social protection efficiency, but the authors frame their tentative effort around the entire welfare system and not the social security field in particular. Total social spending is used as the input variable which includes, amongst other things, public health expenditure and old age pensions, items that do not fit in the state's redistributive effort to reduce poverty and should therefore be kept apart. The analysis is also critically compromised by the use of the post tax-transfer poverty rate and the gini coefficient as alternative output indicators without any consideration for pre tax-transfer income distribution. By doing so, the impact of tax-transfer policies on poverty alleviation is ignored.

Z_{it} is a p×1 vector of variables which may influence the efficiency of a firm; and
δ is an 1×p vector of parameters to be estimated.

Following Battese and Corra (1977), the likelihood function is parameterised in terms of the variance ratio $\gamma=\sigma_U^2/(\sigma_V^2+\sigma_U^2)$. Hence γ indicates the relative magnitude of technical inefficiency variance to total variance in the model.

In estimating efficiency we have used a translogarithmic version of this model[146], which may be described as:

$$\ln(Y_{it}) = \beta_0 \ln(X1_{it}) + \beta_2 \ln(X2_{it}) + \beta_3 \ln(X1_{it})^2 + \beta_4 \ln(X2_{it})^2 + \beta_5 \ln(X1_{it}) \ln(X2_{it}) + (V_{it} - U_{it})$$

Equation 7.5

Y_{it} is the relative reduction of the poverty rate before and after taxes and transfers of country i in period t;
$X1_{it}$ is social expenditure as percentage of GDP of country i in period t;
$X2_{it}$ is the reciprocal of the market poverty rate (before taxes and transfers) of country i in period t;
β and V_{it} are as defined earlier;
t denotes the time period
U_{it} are non-negative random variables are assumed to be independently distributed as truncations at zero of the $N(m_{it},\sigma_U^2)$ distribution, where:

$$m_{it} = \delta_0 + Z1_{it}\delta_1$$

Equation 7.6

$Z1_{it}$ is the targeting of transfers towards the bottom three deciles of the population;
δ_0 and δ_1 are parameters to be estimated.

[146] We use a translog specification due to its flexibility in providing a second order approximation to the underlying production technology.

The measures of technical efficiency relative to the production frontier (4) are defined as:

$$EFF_i = E(Y_i^* \mid U_i, X_i) / E(Y_i^* \mid U_i = 0, X_i)$$

Equation 7.7

7.4 Efficiency Results

The estimation of the stochastic model Equation 7.5 with the inclusion of net targeting as the explanatory variable of the technical inefficiency effects (variable Z1) proved to be problematic. The estimated effect of net targeting on efficiency turned out to be statistically insignificant. Linked to this result, the estimated efficiency frontier was found *not* to be significantly different from the traditional average response function based on ordinary-least-square (OLS) estimation (which suggests all deviations from the production frontier are entirely explained by random noise with no role for inefficiency).

Since there were reasons to be believe that the known unreliability of international comparative data on the size and distribution of direct taxation (Förster, M. and D'Ercole, M., 2005)[147] might be interfering with the estimation procedure, we tested a second specification with the gross targeting of transfers as the explanatory variable of efficiency which provided much better results.

The targeting of transfers was now found to have a significant positive effect on efficiency, which is in line with what the theory requires. Moreover, the stochastic frontier and technical inefficiency effects specification estimated was deemed more appropriate than the traditional average response function based on OLS estimation. All the results presented in this section are based on the estimation of this second specification [Equation 7.5 with gross targeting as the Z1 variable in Equation 7.6][148].

Maximum likelihood estimates of the parameters in the production function frontier and technical inefficiency effects are presented in Table 7.3 and the ensuing efficiency ranking in Table 7.4.

[147] Discrepancies in the treatment of property taxes across different OECD countries are just one example of the comparability limitations of the data on direct taxation provided by the OECD questionnaire on income distribution.

[148] The estimation results of the first specification with net targeting as the Z1 variable are presented in the statistical annex.

Table 7.3: Maximum likelihood estimation results. Absolute t-ratios in parentheses

Production function		Technical inefficiency effects		Variance parameters	
Parameter	Estimate	Parameter	Estimate	Parameter	Estimate
β_0	-4.566	δ_0	-0.756	σ^2	0.031
	(-5.391)		(-1.281)		(3.181)
β_1	-3.126	δ_1	-0.793	γ	0.154
	(-4.183)		(-1.588)		(0.526)
β_2	1.046	Log likelihood	14.915		
	(1.019)				
β_3	-0.559				
	(-2.857)				
β_4	-0.173				
	(-0.332)				
β_5	0.335				
	(0.576)				

Table 7.4: Efficiency ranking

Country	Output Efficiency	Country	Output Efficiency
1995		2000	
United Kingdom	0.991	United kingdom	0.992
Australia	0.988	C.Republic	0.988
New Zealand	0.987	Australia	0.988
C.Republic	0.984	Netherlands	0.987
Netherlands	0.983	New Zealand	0.985
Ireland	0.982	Finland	0.983
Germany	0.979	Denmark	0.980
Sweden	0.977	France	0.978
France	0.975	Sweden	0.977
Finland	0.973	Norway	0.976
United States	0.972	United States	0.973
Denmark	0.971	Switzerland	0.969
Norway	0.969	Japan	0.969
Japan	0.967	Ireland	0.959
Belgium	0.962	Germany	0.958
Canada	0.902	Canada	0.951
Portugal	0.830	Portugal	0.847
Italy	0.721	Italy	0.798

In order to test the robustness of these results we have performed a series of generalized likelihood ratio tests. The likelihood ratio statistic is given by $\lambda = 2(\ln L_1 - \ln L_0)$, where ln L_0 and ln L_1 are the maximum log-likelihood values under the null and alternative hypotheses, H_0 and H_1, respectively. Under the null hypothesis this statistic is usually assumed to be asymptotically distributed as a chi-square random variable with degrees of freedom equal to the number of restrictions involved in the test. However, when the null hypothesis involves a restriction of the type $\gamma=0$ this statistic can be shown to have an asymptotic non-standard mixed qui-square distribution with a degree of freedom equal to the number of parameters restricted to zero under the null hypothesis (Coelli *et. al*, 1998)[149]. The results of the likelihood ratio tests are presented in Table 7.5.

Table 7.5: Likelihood ratio tests (H_{01}, H_{02}: 5% level; H_{03}: 10% level)

Null hypothesis	Λ	Critical Value	Decision
H_{01}: $\delta_1=0$	5.649968	3.84	Reject
H_{02}: $\beta_3=0$; $\beta_4=0$; $\beta_5=0$	17.6347542	7.82	Reject
H_{03}: $\lambda=0$, $\delta_0=0$, $\delta_1=0$	5.649	5.528	Reject

The first null hypothesis H_{01}: $\delta_1=0$, tests the significance of the targeting of transfers in our model assessing the possibility of only $\delta_0\neq0$, which would reduce the model to the panel data version of Aigner, Lovell and Schmidt (1977). The second null hypothesis H_{02}: $\beta_3=0$; $\beta_4=0$; $\beta_5=0$, tests the existence of constant returns to scale applying in a Cobb Douglas model. The third null hypothesis H_{03}: $\lambda=0$, $\delta_0=0$, $\delta_1=0$, tests the appropriateness of the stochastic frontier and technical inefficiency effects specification by assessing the possibility of it being equivalent to the average response function, which can be efficiently estimated by ordinary least squares regression. The three hypotheses are clearly rejected and so we conclude the specification we use to be the most appropriate one.

[149] We obtained the critical value for this statistic from Kodde and Palm (1986).

Overall, the results stress the centrality of the targeting of transfers in explaining the efficiency of social protection systems in reducing poverty. In this respect the efficiency ranking presented in table 3 can be segmented into three main groups:

- The United Kingdom, Australia, New Zealand, the Czech Republic and the Netherlands make up a group of highly efficient countries (both in 1995 and 2000) in reducing poverty through social transfers. The output efficiency scores in all these five countries are above 98 per cent, which means the poverty reduction effect of social transfers is only 2 per cent below the maximum established by our estimated frontier, allowing for each country's individual level of transfers and market income poverty rate. At the basis of such high levels of efficiency are systems of relatively well targeted social transfers (ignoring the effect of direct taxation, the bottom three deciles of the population in these countries receive between 50 and 80 per cent of total non-pension transfers).

- Denmark, Sweden, France, Norway, United States, Switzerland, Japan and Belgium allocate between 40 and 50 per cent of social transfers to the lowest income groups through the course of the two years under analysis achieving an intermediate level of efficiency (output efficiency scores between 96 and 98 per cent, i.e. the poverty reduction effect of social transfers falls short of the frontier by 2-4 per cent)

- Canada, Portugal, and Italy form a third group of relatively inefficient countries where only 25 to 40 per cent of gross social transfers are targeted at the poor and where output efficiency scores are systematically below the 95 per cent threshold (in other words, the poverty reduction effect of social transfers in this group is 95 per cent of the maximum effect determined by the frontier).

As noted above, it has been occasionally argued in the literature that a trade-off exists between the size of transfers and their targeting towards the bottom income tiers of the population. This alleged trade-off is the subject of two contrasting arguments.

The first posits that when social transfers become very large they "tend" to be poorly targeted due to the activity of pressure groups and the progressive universalization of the entitlement to social benefits irrespective of means-testing (Tanzi and Schuknecht, 1997a, 2000). The ensuing policy prescriptions of this

literature are the confinement of the size of social transfers within moderate levels and the concentration of efforts in improving the targeting of transfers.

Our results do not corroborate this thesis. Both the correlation between the size of social transfers and their degree of targeting towards the poor as well as the correlation between the size of social transfers and the efficiency scores of Table 7.4 are non significant[150]. Hence, according to these results size is not linked to targeting or to efficiency - there is no trade-off[151].

The second argument asserts that the size-targeting trade-off stems from the political process. In the long run focusing on the improvement of the targeting of transfers is said to be likely to reduce the total amount of resources available for redistribution and to have an overall negative impact on poverty alleviation[152] (Korpi, 1983; Korpi and Palme, 1998; Mahler and Jesuit, 2004; Nelson, 2004). Baldwin (1990, p. 298) provides an interesting account of the alleged phenomenon:

***

…In nations where the state became the main insurance broker of the bourgeoisie, in contrast, the disadvantaged gained from clinging to the coattails of the favoured. The middle classes arranged things first and foremost for themselves, the unfortunate were beneficiaries of a comparatively successful trickle-down…In the long run, the unfortunate have gained most from those welfare states securely anchored in the interests and affections of the bourgeoisie.

***

Hence, the size of highly targeted programs is thought to be limited by the lack of political power of their main supporters (the poor), whereas social programs with extensive coverage are expected to capture support from a much more powerful political base (Korpi, 1983). Following this line of reasoning, granting high levels of income security to middle and high-income groups is supposed to increase the possibilities of also providing income protection for the poor[153].

---

[150] The pearson correlation coefficient between efficiency and social expenditure is 0.186 and not statistically significant. The same coefficient for social expenditure and targeting is 0.185 and also non statistically significant.

[151] In fact, all the countries we identified as being relatively inefficient in redistributing income (with low levels of targeting towards the poor) have relatively small social security budgets when expressed as a percentage of GDP.

[152] Korpi and Palme (1998, p.663) have designated this as the "paradox of redistribution" whereby "the more we target benefits to the poor…the less likely we are to reduce poverty and inequality".

[153] An argument that has also become to be known as the middle-class inclusion thesis (Pedersen, 1999).

Our results do not corroborate this argument either. Firstly, has it has already been stated our evidence does not support the notion of a size-targeting trade-off. Secondly, we find no evidence of a link between the targeting of transfers and poverty alleviation (measured through our output variable)[154]. Therefore, the idea that better targeted transfers have an overall negative impact on poverty alleviation due to the preponderance of an indirect negative effect caused by an alleged reduction in redistributive funds is not supported by our dataset.

A common thread to both of these arguments is their reliance on anecdotal empirical evidence of a purely descriptive nature and merely tentative efforts at measuring the concepts involved.

Our study differs from this literature both in terms of the methods employed in the examination of the states' redistributive policies and the analytical results derived from that examination. We have used refined input and output indicators that purged non-poverty-alleviation elements from the analysis[155]; the efficiency of social transfers has been assed in accordance with a formal efficiency analysis framework; and the study has been conducted for a sample larger than that of any previous work. The results thus obtained underline the importance of the targeting of transfers as the key variable for explaining the efficiency of social transfers and do not corroborate the existence of an empirical size-targeting trade-off.

7.5 Modes of Governance and Social Protection Efficiency

After measuring efficiency and tracing it back to the targeting of transfers, the subsequent logical step is to investigate the institutional factors that lead different countries to achieve different targeting patterns. In order to do so we need to allow for the significant variety of governance arrangements that characterise modern social security systems.

The literature on the welfare state is permeated by organisational typologies. The most influential of these is probably Esping Andersen's (1990) classification of

[154] The pearson correlation coefficient between these two variables is 0.227 and non statistically significant.
[155] The input indicator does not include health expenditure or old-age pensions. The output indicator reflects the impact of tax-transfers policies specifically on poverty alleviation and not on the entire distribution of income (as in some of the previous works).

welfare regimes, which became a source of inspiration for most of the categorisation efforts that followed (e.g. Lewis, 1992; Castles and Mitchell, 1993; Ferrera, 1996, 1998; and Korpi and Palme, 1998), and is itself inspired by the tradition of Titmuss (1958, 1974). However, our use of these typologies to explain the targeting of transfers and efficiency is compromised both by the prominent role old age pensions play in them as well as by their reliance on political/institutional factors that escape the scope of our analysis (such as political orientation and the influence of labour unions).

The classification developed by Kraus (2004) enables us to overcome these problems by limiting the analysis to monetary social transfers other than pensions and by focusing specifically on the characteristics of social transfers systems that directly relate to the social security strategies employed.

The author identifies three ideal-type social strategies[156] in Europe. The *social assistance strategy* aims at poverty mitigation through the provision of a socially acceptable minimum support for those with inadequate income for reasons of misfortune. The *social insurance strategy* focuses on poverty prevention through the provision of income replacement benefits as part of an underlying compulsory insurance contract. This strategy has been adopted under two different approaches. In the *Bismarck-type social insurance* coverage is limited to specific groups (normally defined by occupational category), and benefit entitlement is dependent and related to past contributions/earnings. In the *Beveridge-type social insurance* benefit entitlement is also dependent on past contributions but benefits are generally flat-rate as the system seeks to reach a more comprehensive coverage. Finally, the *social allowance strategy* is based on the idea of universal guaranteed minimum income as a right of nation-state citizenship. It foresees the concession of grants whose entitlement and level does not depend on neediness, past earnings or contributions[157].

It is argued that real-world social security systems do not conform to any single ideal-type social strategy but instead comprise a mixture of different strategies, the relative importance of each strategy varying from one system to the other. In order to examine the differences between European social transfer systems in terms of the adopted social security strategy mixes, the author develops a classification of social

[156] Drawing on the work of Esping Andersen (1990), Hill (1996) and Dixon (1999).
[157] The most universal social allowance would consist in the granting of some basic income to every citizen regardless of any criteria other than citizenship (Hill, 1996). In practice it is usually adopted some form of demographic criteria such as number of children and age.

transfer systems according to the extent to which the various strategies are employed. The three key distinctive features considered are the funding regime, benefit levels and benefit entitlement. The following pairs of indicators were taken as proxies of these key features (Kraus, 2004, p.437)[158]:

- With respect to funding issues the share of transfers in GDP and the ratio of funding by the state to funding by contributions;

- With respect to the level of protection the ratio of minimum income guaranteed to median equivalsed income for single adults and an indicator for income replacement rates of social transfers;

- With respect to the conditions for benefit entitlement the share of means tested benefits in social expenditures and an indicator for the degree of coverage of social transfers.

By applying cluster analysis to this set of indicators the author concludes that within the European social transfer systems four general types of regimes can be identified (Kraus 2004, p.442, 443):

- The *Southern European Cluster* (Greece, Portugal, and Italy) where the dominant strategy adopted is Bismarck-type social insurance only marginally supplemented by additional measures of social assistance or allowances. Overall coverage is fragmented, means tested assistance is of minor importance and social allowances are either inexistent or play an insignificant residual role;

- The *Central European Cluster* (Germany, Belgium, Austria, and France) where the dominant strategy is Bismarck-type social insurance supplemented by social assistance measures to provide a guaranteed minimum income for those not covered by the main strategy and a small amount of social allowances, particularly in the field of family benefits;

- The *Northern European Cluster* (Finland, Sweden, Denmark, and the Netherlands) where the dominant strategy can be described as Beveridge-type social insurance supplemented to a relatively large extent by non-contributory-based social allowances. Overall coverage is the highest in Europe, and means tested social assistance varies in importance between countries;

[158] The underlying dataset and the subsequent analysis is for the mid-1990s.

- The *British Cluster* (UK and Ireland) where the dominant strategy is Beveridge-type social insurance supplemented by social assistance with means tests playing a prominent role and a high degree of coverage.

This class assignment of social security regimes seen in the light of the efficiency ranking we obtained for the mid-1990s (section 7.4, Table 7.4) yields some insightful results. The members of the British cluster[159] are unequivocally the best performers in terms of efficiency with average output efficiency scores of 98%. In sharp contrast, the Southern European Cluster stands out by its clear concentration at the bottom of the table (average efficiency 77%). In between these extremes, the Northern and Central clusters occupy a space that is difficult to disentangle (the average efficiency score is of 97% for both clusters[160]). Kraus found similar results when the same clusters were associated with the redistributive effects of social transfers[161]:

***

"The Beveridge-con-social-assistance strategy of the British cluster leads to a redistributive pattern that clearly reflects the simultaneous impact of both the poor law tradition [means tested social assistance] and the Beveridge-goal of universal coverage. Consequently, the system is highly favourable to lower income groups, whereas recipients of medium and high incomes are markedly less favoured…the Bismarck-con-social-assistance strategy of the Central cluster is primarily directed towards the protection of workers and places more emphasis on horizontal redistribution. Accordingly, medium income levels benefit to a comparatively large degree, while the low income groups are still covered by minimum income guarantees. As the latter are all but absent in the Southern cluster, the Bismarck strategy adopted produces only minor vertical redistribution. Finally, with regard to the Northern cluster, the Beveridge-con-social-allowance strategy produces a redistributive pattern half-way between the Central and the British cluster, being more favourable to low income groups than the Central cluster, whereas high income groups benefit stronger than in the British group".

Kraus (2004, p.451, 454)

***

[159] Apart from the United Kingdom and Ireland it may be reasonable to consider additionally Australia and New Zealand as being part of this cluster given the evident similarities in terms of policy and social security indicators between these countries. However, in the absence of further research, the inclusion of these countries in the British cluster remains essentially speculative.

[160] We are considering Denmark as making part of the Northern Cluster again for reasons of similarity in terms of social security policies in relation to the other members of the group. Again, the same note of caution on the validity of such judgement applies as in the extension of the British cluster.

[161] Measured by the skweness and kurtosis of the distribution by the different quintiles of the population of the increase of average income due to social transfers.

Our results show that when the effect of different social security regimes is assessed not on the overall redistributive pattern they create but rather on the efficiency of social transfers in alleviating poverty the relative position of the Northern and Central clusters becomes ambiguous. This may be explained by the simultaneous presence of two competing forces. One the one hand, the Beveridge-type social transfers of the Northern cluster are expected to produce better results in terms of efficiency when compared with the intrinsically less well targeted transfers of Bismarckian systems[162]. On the other hand, the social allowances schemes characteristic of the Northern cluster are expected to be detrimental to efficiency when compared with the supplementary means tested social assistance of the central cluster. The relative weight of these two counteracting forces ultimately determines which of the two models is linked to higher levels of efficiency. The short number of observations in our sample does not allow us to draw any definite conclusions on that.

The results also show that the two contrasting policy approaches of the British and Southern clusters yield equally contrasting efficiency results. This can be explained by differences both at the level of the main social security strategy being followed as well as of the supplementary schemes adopted.

Bismarckian systems place an emphasis on the role of the state as a monopolistic insurance provider for given sectors of the population. With a view to protecting individuals belonging to certain corporatist groups from inappropriate voluntary self-protection[163], the state takes full responsibility for providing compulsory insurance for most of the relevant social risks, leaving a very small margin for individual voluntary action in the insurance market. The result is that a substantial share of social transfers is allocated on a categorical basis (without any means test), leading to substantial horizontal redistribution[164]. The targeting of social transfers is generally poor, and thus the system is deemed inefficient according to our measurement criteria.

Beveridge-type systems are instead informed by the notion of a basic role for the state in social insurance provision. Although it is explicitly acknowledge the need

[162] Flat-rate benefits of typical Beveridge-type schemes tend to create a redistributive pattern that is relatively favourable to the low income deciles of the population.
[163] Which may result from individual short-sightedness as well as problems of adverse selection and/or imperfect information.
[164] For example from active to currently non-active persons; from families without children to families with numerous children.

for state intervention in the form of flat-rate insurance benefits covering as large a proportion of the population as possible, it is simultaneously recognised that any level of protection above the minimum is expected to be acquired in the market through voluntary social insurance. This policy approach clearly improves the targeting of transfers and consequently the efficiency of social expenditure.

These two policy strategies epitomise two radically different approaches to the governance of social security. The first prescribes a bureaucratic, centrally-driven remedy for dealing with a market failure. The second foresees the need for some level of central coordination in assuring a minimum level of insurance provision, but leaves the rest of the system to be coordinated by decentralised decision making procedures.

Apart from the insurance component an equally important part of social security systems is to do with purely redistributive expenditure intended to assist individuals in need that did not make significant contributions in the past entitling them to social insurance. Means tested social assistance is the policy instrument that overtly attempts to fulfil that role[165]. Its residual role in the Southern cluster also helps to explain why in this case we found evidence of relative inefficiency.

7.6 Conclusion

Benefit systems redistribute income. But they do not primarily redistribute from rich to poor. Rather, they redistribute from young to old, from those who work to those who do not, and from childless families to families with children. In most countries benefits are not so much based on the income of the individual or family, but on the circumstances of the family and the individuals that make up the family more generally (Forster and Pearson, 2002).

In the mid-1990s, on average the bottom 30 per cent of the population received 36,16% of total benefit payments. The top 30 per cent of the population got an average of 25,89% of all benefits. The middle 40% of the population absorbed the remaining 37,95%. In other words, the benefit system did not have a very different effect on final income inequality from paying everyone in the population a fixed amount of benefit, regardless of income level.

[165] In this respect it must be said that universal social allowances are neither intended nor able to fulfil that role.

On the basis of such poor targeting are two key factors. The first is the progressive transformation of social security systems from mechanisms of collective insurance, providing services that are not properly supplied by the market, into instruments for the pursuit of obscure political agendas of equality. The universalization of benefits (e.g. social allowances) is the practical manifestation of such transformation, and it collides quite evidently with the initial theoretical drive of social security (and indeed the only one that can be justified on economic grounds): poverty alleviation, be it as a direct counterpart to the contributions made by individuals when they were not in poverty (insurance component), be it as gesture of social solidarity when the recipient has not contributed to the system and would thus otherwise be left uninsured (purely redistributive component). In practice, the universalization of benefits enables the middle and richest classes to get hold of resources that should either be channelled to the poorest deciles of the population or that should not have been appropriated by the state in the first place.

The second factor is fiscal churning. Palda (1997, p. 190) defines a transfer as being churned "when the person who receives it would have been just as well off, or better off, with a tax cut of the same size as that of the transfer". Money is taxed out of a citizen's pocket, filtered through a government bureaucracy, and sent back into that same citizen's pocket. This process needlessly destroys resources: the tax and subsidy misdirect the individual's economic efforts and the bureaucracy that overlooks transfers and taxes consumes resources without any productive counterpart.

Chapter 8: Conclusion

The size and role of government has long been the subject of controversy. It is plain that market economies cannot function without government. It is equally clear there are limits to the extent that it is desirable to expand government activity. The optimal balance between market and government remains an open debate.

There is ample evidence of the potential of government expenditure on a core set of activities to improve economic performance by fostering the accumulation of human and physical capital, raising the rate of technical progress, and creating synergies with private activities[166]. On the other hand, there is equally ample evidence that as government outgrows this set of core activities it can crowd out more productive private sector activity and depress economic performance. This effect is further amplified by the reliance on distortional taxation to finance growing public sector activity, which imposes allocation effects over and above those related to the level and composition of public expenditure[167].

Therefore, as government expands its activities there must come a point where social welfare would increase if people had been allowed to keep their money to spend for themselves instead of paying extra taxes to finance expansionary public services. In theory at least, there must be an optimal *level* (and *structure*) of taxation to finance an optimal *level* of *core* public services provided in an *efficient* way.

The political science and public management literature is filled with analytical typologies on the governance of public services, but fails to establish a link with efficiency. The economics literature, in contrast, provides useful tools for understanding the nature of the relationship between governance and efficiency, but it still lacks a unified body of theory and it is sill at its infancy when it comes to measuring public service efficiency and investigating its causes.

NIE (arguably the most influential economic theory of organisation) lays emphasis on the role of transaction costs arising from contractual hazards associated with the propensity of individuals to behave opportunistically and noncontractible environmental uncertainties. Costs of this kind are important (even if not the most important) to explain the boundaries of organisations (and in particular firms'

[166] The actual level of expenditure depends on the level of efficiency of government activities.
[167] Most taxes distort economic behaviour by influencing people's preferences between consumption (work) and leisure, or by interfering with the incentives to invest or save. Expansionary public sector activities financed through distortional taxation imply an increase in taxes rates that is more than proportional to the required addition on tax revenue. The excess burden or deadweight lost associated with taxation increases with the square of the marginal tax rate (Dupuit, 1844).

decisions to "make or buy" inputs), but the claim they explain the existence of firms reduces the role of conscious central coordination to vanishing point.

Resource-based/evolutionary perspectives of the firm bring the limitations of NIE to the fore by drawing our attention to the role of interrelationships between activities, and associations of capabilities that support them in explaining the need for hierarchical coordination.

Although NIE and resource-based theories of organisation complement each other in a certain way, we found both to be insufficient to grasp the full extent of the link between economic organisation and efficiency, not least because of their exclusive focus on the organisation of production.

Following Hayek's seminal distinction between *spontaneous self generating order* and *organized ("made") order* we have argued that economic organisation is essentially about bringing order into systems of human interaction. Order can be accomplished through a multitude of modes of governance, ranging from conscious deliberate design to spontaneous self-reinforcing rules. Running each of these modes of governance entails costs of coordination and motivation of individual decisions which vary with the attributes of the transactions[168] upon which they are deployed. Minimising these costs is the main problem of economic organisation.

No system can coordinate overall economic activity more efficiently than a system of markets (Arrow and Debreu, 1954). The informational requirements of the price mechanism are minimal when compared with others systems capable of producing an equally effective outcome.

However, decentralised coordination of transactions for which there is a great deal of *a priori* knowledge about the optimal solution (i.e. with *design* attributes - Milgrom and Roberts, 1992) is inefficient due to the brittleness of the system in dealing with imperfect information and excessive demands of communication. Central coordination through formal organisations structured along hierarchical lines is, in this case, optimal.

Decentralised coordination is also sub-optimal when transactions exhibit public good attributes, externalities, merit good attributes, and/or information asymmetries. *Public* central coordination is, in this case, optimal.

[168] Some authors (e.g. Milgrom and Roberts, 1992) refer to these cost as "transaction costs", in the sense that they represent the costs of coordinating economic transactions. However, these should not be mistaken for the typical transaction costs alluded to by TCE, which make up only a small part of the former.

From the need to correct the market does not follow immediately, or necessarily, the need, or desirability, to fully replace it. We have argued in this work that while government should interfere with specific attributes of the organisation of services that markets fail to address, it will still be desirable to rely on market forces as thoroughly as possible for the remaining attributes of services they coordinate efficiently. What is more, government corrective interference should openly acknowledge the limits of central coordination and associated intrinsic risks of government failure.

The key policy implication that follows is that the detailed design conditions of the organisation of public services need to be rigorously tailored to the specific characteristics of each service. Depending on these characteristics, the optimal mode of organisation will lie within a continuum of funding and provision arrangements, ranging from small scale regulation of private sector activities to universal monopolistic provision of services through centralised public bureaucracies.

In the private economy, profit-maximising producers and utility maximising consumers harmonise their individual goals using the price mechanism. In a perfectly competitive market, prices adjust to that ensure supply equals demand for each good and service, and that the allocation of resources is Pareto efficient - given all the restrictions in terms of available resources and technological possibilities, no other allocation is unanimously preferred.

In the "public" economy, using prices in the form of fees/charges as a means of paying for the services citizens consume is also associated with a number of positive outcomes[169]. However, often some of the specific attributes of services that markets fail to address are precisely to do with the funding side of the service and require public intervention.

Education offers one of the most compelling cases for public funding with positive efficiency implications. Capital market imperfections are more likely to affect investment in human than in physical capital (as the later can serve as collateral), forcing individuals to underinvest in education. Education is also known to have important effects on citizen involvement in the community and government,

[169] Charging consumers a price equal to the marginal cost of provision is needed to induce *both* consumers and providers to behave in ways that ensure efficiency; political accountability is strengthened through strong and clear awareness of the costs and benefits of services made to consumers; the burden of distortionary direct taxation and/or regressive indirect taxation is eased by funding services through fees/charges.

crime, family decision making, child upbringing, and economic growth. Public funding of education is required to allow for these credit-constraints and external effects, pushing the rate of human capital accumulation closer to its optimal level.

In healthcare, higher propensity of those with greater health risks to buy insurance and to insure at higher levels than those in good health limits access to affordable insurance for high risk individuals, lowers coverage and leads to under-consumption of healthcare from a social optimum point of view (Docteur and Oxley, 2003). Public funding is required to make the insurance compulsorily, regulate the ability of individuals and organisations to accept/reject contracts, and ensuring that a pooling equilibrium is reached. Chapter 6 provided evidence on the inefficiency of short-term market interactions between patients and providers, or unregulated private insurance systems as mechanisms for funding healthcare.

Similarly to healthcare, moral hazard and adverse selection make public funding of social protection systems necessary to pool social risks effectively. Unlike healthcare, though, public funding of insurance for social risks is additionally required to deal with problems created by uncertainty about the size of future claims and risk correlation in time.

When market failures are confined to the funding attributes of the service (e.g. transport), private provision is enough to ensure efficiency (provided that the other necessary conditions for effective competition are in place). When market failures extend to the provision side of the service (e.g. education and healthcare), private provision alone does not ensure effective competition, or efficiency.

Market failures at the service provision level generally stem from: (i) information asymmetries that cloud the ability of consumers (and government) to effectively assess the quantity and quality of service provided; (ii) market power following natural monopolistic/oligopolistic market structures at the regional/local level. Any of these two failures, or a combination of both, require the public sector to play an active role in the provision of the service, either by replacing (e.g. police), supplementing (e.g. health care and education), or regulating private provision (e.g. transport)[170]. When the public sector replaces or supplements the private sector, the

[170] Other institutional arrangements typically elicited to address these market failures include professional norms, licensure, and non-profit organisations.

resulting organisational arrangement is inevitably sub-optimal when compared with the perfect competitive market theorised by neoclassical economists[171].

Moderate information asymmetries combined with a monopolistic/oligopolistic market structure[172], call for *partial* public sector involvement in the provision of the service. Public providers are likely to counteract the potential perverse effects of combining information imperfections with private providers' incentives to exploit market power. Private providers, in turn, are likely to sharpen the incentives of public providers to behave efficiently. Primary and secondary education is a good example of such type of service, and our empirical results in chapter 5 appropriately demonstrated that (i) efficiency declined as the share of public providers increased; and (ii) efficiency improved with the degree of decentralisation of decision making processes within publicly owned and managed schools (i.e. with their degree of managerial autonomy).

When information asymmetries are particularly severe, limiting substantially the ability of consumers and/or government to monitor the quantity and quality of service outputs, the case for involving the private sector in the provision of the service becomes very weak. In such case, the public sector is then left with the task of fully replacing the market. Medical care is a good example of this type of service, and our empirical results in chapter 6 showed evidence of efficiency being higher in public integrated systems (typically consisting of a majority of public providers) as opposed to systems of private (for-profit and non-for-profit) provision.

Finally, our study of the governance and efficiency of social protection systems took us through a distinct analytical path. Private social insurance markets are prone to failure to such an extent that, even in theory, they cannot be conceived of as a viable mode of governance for social protection. Here, the choice is thus not between public or private provision of social protection, but merely greater or lesser public involvement. Accordingly, what our hypothesis suggested, and our empirical results confirmed, was that as public social security systems exceed their remit to

[171] Even if we assume the existence of a performance management system that renders effective incentives for managers of public organisations to maximise efficiency in a given moment, there are likely to be obstacles attached to public ownership that will eventually damage efficiency in the medium or long run. A typical example would be managers spotting an opportunity for making an investment in a particular source of innovation, but not being able to obtain from the public sector the financial support required to make that investment, or not being able to capture adequate rewards for having done so.

[172] As a result of economies of scale/scope in the production of the service together with its provision at the local/regional level.

assist individuals with the process of smoothing income (across the life cycle and across states of nature), and to provide social safety nets (to those whose total income over the entire life-cycle is insufficient to sustain an adequate living standard), efficiency is damaged. We also found that efficiency is not affected by the level of social protection provided within the remit defined by our hypothesis. This contradicts other authors' suggestion of the existence of a trade-off between the size of transfers and their targeting towards the bottom income tiers of the population – either due to the activity of pressure groups and the progressive universalization of the entitlement to social benefits irrespective of means-testing; or because granting high levels of income security to middle and high-income groups *allegedly* increases the political support for redistribution in a way that improves the possibilities of providing income protection for the poor.

This thesis has made significant progress in understanding the organisational drivers of efficiency. Nevertheless, there is still a long way to go in the quest for understanding the key levers that affect the efficiency of public services. There are several areas where further research is likely to yield results of utmost importance.

First, much is to be expected from exploring in greater depth the attributes of transactions that call for central coordination. In particular, the development of a conceptual framework, similar to the one used by TCE, that refines the study of the interaction of public good attributes, externalities, and information asymmetries with different modes of governance to the level of individual services would certainly shed light on elements that are currently overlooked by the macro scale of our analysis.

In the education sector, there are clear gains to be made from better understanding the theoretical and practical implications of the ownership of schools (especially between publicly owned, privately managed schools on the one hand, and privately owned and managed schools on the other) on efficiency, in the context imperfect information on school outputs.

In the health sector, the current reliance on disability-adjusted life expectancy and infant mortality rates as outcome measures of health systems at the international level is regrettable not only because of the inability of these variables to capture thoroughly the health status of individuals, but also because they fail to account for non-medical determinants of health (e.g. physical exercise; dietary habits, etc) whose

effects cannot be ignored (and for which there is currently no cross-country data available).

In the social protection sector, additional effort should go into unveiling the relationship between the size of social transfers and the extent to which they are targeted towards the lowest income classes. Although we did not find evidence of a significant correlation between the two, we did notice that none of the most efficient countries were high spenders, and there appeared to be signs of inefficiency when high spending countries such as Sweden, Finland, Denmark, or Netherlands were compared with countries like the United Kingdom, Australia, New Zealand or the Czech Republic. One could argue that more spending leads to less targeting and greater inefficiency as it both enables and encourages rent seeking behaviour of interest groups. It may also be argued that inefficiency associated with churned transfers is conducive to higher social spending[173]. The theoretical link between size, targeting, and efficiency needs further investigation.

Second, in the near future it is unlikely that significant efficiency improvements will arise from radical shifts in the organisation of public services. The potential for improvement lies in fine tuning the role the public sector plays in the governance of these services. A substantial part of this process will involve forging new hybrid governance arrangements both within public sector and in the way it interacts with the private and voluntary sectors. There is already a considerable body of theory focusing on the inner workings of these hybrid modes of governance, but any sensible map of future lines of research must include incentives to move ahead with this line of investigation.

Third, the measurement of public service efficiency is an exceptionally complex methodological task that still faces enormous obstacles. The determination of the societal value of public services and aggregation of public service outputs is central to the study of the optimal forms organising public services. We need better measures of the output of the public sector than the ones currently available at the international level to explore fully the learning potential engrained in the rich diversity of governance structures deployed by different countries to similar services. Equally, there needs to be a better understanding of the relative preferences of society

[173] It is doubtful whether the governments of high spending countries would be able to collect the amount of taxes they do, if a significant portion of those resources were not channelled back again to their original source.

for public services if we are to be able to monitor and influence the effectiveness and efficiency of government activity in the future.

The importance of this point cannot be overstated. If information imperfections are extreme, or the quality of available measures of public sector output is very poor, in the sense that the later is essentially undetermined, replacing a failing market with public provision may improve service effectiveness but does not necessarily ensure efficiency[174]. Conversely, in a scenario of non-competitive market structures (monopoly/oligopoly) and perfect information, private provision combined with price regulation[175] is likely to bring about an efficient outcome[176].

Monitoring and managing the performance of public services appropriately implies using information on inputs, outputs and environmental constraints for all relevant providers to estimate efficient costs of provision, which can then be drawn upon to identify good/poor performance and trigger changes: e.g. inefficient providers losing market share; managerial boards of inefficient providers being subject to sanctions; expanding existing efficient providers; and allowing new providers to operate. The benefits of such a relative performance management exercise will arguably be improved if users are able to choose between providers, but they do not critically rely on the availability of choice, nor should they be confused with the benefits of choice (in the cases where indeed there are some benefits associated with choice).

The potential gains to having the public sector involved in the provision of a given service (alone or together with private providers) will only be fully explored if public provision is, in itself, subject to competitive pressures that diffuse incentives for efficient behaviour. In order for that to happen, information on the relative performance of public sector organisations needs to be available and used to elicit corrective action.

---

[174] The lack of a profit-motive for public providers may lead them to deliver higher levels of service outputs than profit-seeking providers knowledgeable of the difficulties of individuals (or government) to monitor their performance. Conversely, it is equally reasonable to assume that the profit-motive will lead private providers to be technically more efficient than public providers not subject to any kind of proper performance monitoring.

[175] E.g. the yardstick competition model proprosed by Schleifer (1985).

[176] Kenneth Arrow's (2001, p. 1202) vision of the future role of the public sector in this area is particularly insightful: "I do believe we will see a transformation of the economy and of the medical care market because of the information revolution. I think it can take many decades to really understand how to use our information powers. It may well be that there will still be asymmetry of information, in the form of a group that certifies the information, even at the end of the process".

When the nature of services is such that requires them to be provided by regional/local public monopolies (e.g. police services), relative performance management is likely to be particularly valuable. By using information on the costs of the various providers to infer a benchmark against which to evaluate individual organisational performance, government can infuse competitive pressures in a system that, otherwise, lacks strong incentives to behave efficiently.

Local government activity is an area where a variant of this approach could also yield important results (Besley and Case, 1995; Bordignon, Cerniglia and Revelli, 2002). In the presence of imperfect information regarding the behaviour of local officials, the local electorate can look at the performance of other (neighbouring) jurisdictions to provide a benchmark against which to assess the performance of their own elected officials.

Measuring activities at a refined level of detail has already proven to be possible across a wide range of health care and education services in some countries (a good example is the UK). Significant efforts are currently being developed by EU members to improve the measurement of non-market output in national accounts (Smith and Street, 2006). A persistent concern remains, though, as to how the various activities or outputs should be valued. Traditionally, costs have been used as a proxy for value, which effectively assumes they reflect closely marginal benefits (something that in a non-market setting is very unlikely). Ultimately, the selection of objectives and the determination of their weights is the responsibility of politicians who are charged with reconciling conflicting claims on public resources.

What public services are to be provided, by whom and in what way, are questions that lie at the heart of electoral politics in contemporary democracies, and are likely to continue to do so. The questions are much the same in all countries, yet the answers adopted differ considerably. One explanation is that history matters. The organisation of public services is shaped by dynamic relations with some degree of instability, so that small variations can produce large and long-lasting deviations in outcome overtime. Structures are costly to change, and as soon as social norms and political decisions arise, the governance arrangements tend to exhibit high sensitivity to initial conditions (Arrow, 2001).

The precise nature of the social contract implicit in the funding of public services provides a good illustration of such path dependencies. While most countries

have made the financing of education and healthcare the responsibility of the state, some (most notably the United States) have left it to the responsibility of families, local communities, and employers. This diversity in policy approaches is not necessarily explained by discrepancies in tastes, technologies, or political systems. A plausible explanation is simply that history matters. Temporary shocks to the distribution of wealth as well as to the political system are known to be capable of permanently moving society from one equilibrium to another, with long-lasting effects on the economy and the social contract.

In a path dependant policy world, where governance arrangements are often as much the product of careful policy design as of circumstantial determination, understanding the processes that work to shape the efficiency of public services can yield an invaluable contribution to policy making. Over the last decades, major changes have been attempted or advocated in various public services across a range of jurisdictions. The boundaries between the public and private sector have shifted in a number of areas and are likely to continue to do so. New modes of delivery and new provider structures have been tested and adapted to the political and institutional context in which public services operate. This thesis has drawn on the experience of various countries operating similar public services under different organisational regimes to shed light on the policy drivers of public service efficiency. The next fundamental step is to combine the theory and evidence presented in this work with the current knowledge on the optimal levels and structures of taxation and public expenditure to achieve a better understanding of the optimal size and role of State in modern societies.

Statistical Annex – Data and Sources

Table 6: Education data and sources

Country	Pisa Average Score (1)	Cumulative expenditure p/std (2)	Index of economic social and cultural status (3)	%Decision-making at school and local levels (4)	%Public providers (5)
Austria (2003)	496	73329	0.06	0.51	0.92
C.Republic (2003)	511	23144	0.16	0.92	0.93
Denmark (2003)	493	67037	0.20	0.81	0.78
Finland (2003)	544	50645	0.25	0.98	0.93
Germany (2003)	499	47660	0.16	0.49	0.92
Greece (2003)	463	28791	-0.15	0.16	0.97
Hungary (2003)	491	18190	-0.07	0.96	0.89
Iceland (2003)	498	54144	0.69	0.75	1.00
Italy (2003)	474	68939	-0.11	0.63	0.96
Japan (2003)	530	63351	-0.08	0.67	0.73
Korea (2003)	541	39017	-0.10	0.56	0.42
Mexico (2003)	393	10480	-1.13	0.22	0.87
Netherlands (2003)	525	48394	0.10	1.00	0.23
Norway (2003)	490	60193	0.61	0.68	0.99
Portugal (2003)	468	42048	-0.63	0.41	0.94
S.Republic (2003)	488	11893	-0.08	0.65	0.87
Spain (2003)	482	40424	-0.30	0.28	0.64
Sweden (2003)	507	56668	0.25	0.83	0.96
Austria (2000)	510	71387	0.01	0.47	0.89
C.Republic (2000)	500	21384	-0.04	0.62	0.94
Denmark (2000)	497	65794	0.20	0.74	0.76
Finland (2000)	538	45363	0.04	1.00	0.97
Germany (2000)	485	41978	0.16	0.53	0.96
Greece (2000)	458	27356	-0.08	0.23	0.96
Hungary (2000)	485	20277	-0.05	1.00	0.95
Italy (2000)	471	60824	-0.17	0.36	0.94
Korea (2000)	537	30844	-0.17	0.32	0.51
Norway (2000)	499	61677	0.49	0.65	0.99
Portugal (2000)	458	36521	-0.58	0.24	0.93
Spain (2000)	485	36699	-0.39	0.41	0.62

(1) Arithmetic average of PISA Mathematics (space and change; change and relationships), Reading and Science scores (OECD, 2004b, Tables 2.1c; 2.1d; 2.2c; 2.2d; 6.2; 6.6; OECD, 2001. Tables 3.3;2.3a;3.6).
(2) Cumulative expenditure per student between 6 and 15 years-old in purchasing power parities (OECD, 2004b, Table 2.6 (values deflated to 1998); OECD, 2001, Table 3.6).
(3) Mean Pisa Index of Economic and Social Status (OECD, 2004b, Tables 2.6;4.3b).
(4) Percentage of decisions relating to public sector, lower secondary education, taken at each level of government (OECD, 2004a, Table D6.6).
(5) Percentage of students enrolled in Public Schools (OECD 2004b, Table 5.19; OECD, 2001, Table 7.13).

Table 7: Social protection data and sources

Country	Poverty rate, market income (1)	Poverty rate, disposable income (2)	Effect of taxes and transfers in reducing poverty (3)	Output Variable (4)=(3)/(1)	Public non-pension cash benefits (%GDP): Input Variable (5)	Gross Targeting (6)	Churning ($taxes_{poor}/transf_{poor}$) (7)	Net targeting (8)=(6)*[1-(7)]
AUS (mid-1990s)	0.205	0.075	0.130	0.635	0.059	0.621	0.122	0.545
CAN (mid-1990s)	0.178	0.094	0.084	0.473	0.040	0.382	0.468	0.203
C.REP (1992)	0.177	0.030	0.146	0.828	0.062	0.422	m	m
DEN (mid-1990s)	0.205	0.045	0.160	0.780	0.110	0.456	0.559	0.201
FIN (mid-1990s)	0.181	0.054	0.127	0.702	0.128	0.432	0.392	0.263
FRA (mid-1990s)	0.261	0.068	0.194	0.741	0.079	0.356	0.121	0.313
GER (mid-1990s)	0.186	0.072	0.113	0.611	0.056	0.388	0.698	0.117
IRL (mid-1990s)	0.266	0.086	0.180	0.676	0.090	0.471	0.094	0.427
ITA (mid-1990s)	0.236	0.127	0.109	0.463	0.064	0.168	0.611	0.065
JPN (mid-1990s)	0.140	0.119	0.022	0.155	0.023	0.293	0.942	0.017
NLD (mid-1990s)	0.176	0.062	0.114	0.649	0.110	0.458	0.603	0.182
N.ZLD (mid-1990s)	0.182	0.070	0.112	0.616	0.068	0.547	0.344	0.358
NOR (mid-1990s)	0.142	0.067	0.075	0.530	0.081	0.451	0.522	0.215
POR (mid-1990s)	0.166	0.100	0.066	0.400	0.053	0.217	0.351	0.141
SWE (mid-1990s)	0.186	0.042	0.145	0.776	0.092	0.383	0.560	0.168
UK (mid-1990s)	0.204	0.080	0.125	0.610	0.078	0.545	0.225	0.422
USA (mid-1990s)	0.187	0.135	0.052	0.278	0.031	0.414	0.498	0.208
AUS (2000)	0.205	0.086	0.119	0.580	0.058	0.628	0.141	0.539
CAN (2000)	0.160	0.103	0.057	0.356	0.030	0.394	0.594	0.160
C.REP (1996)	0.195	0.038	0.157	0.807	0.057	0.453	0.274	0.329
DEN (2000)	0.185	0.050	0.135	0.731	0.086	0.506	0.591	0.207
FIN (2000)	0.153	0.064	0.088	0.578	0.085	0.501	0.459	0.271
FRA (2000)	0.241	0.060	0.181	0.749	0.075	0.363	0.139	0.313
GER (2000)	0.205	0.080	0.125	0.610	0.048	0.335	0.638	0.122
IRL (2000)	0.188	0.119	0.069	0.368	0.053	0.455	0.194	0.367
ITA (2000)	0.218	0.115	0.103	0.472	0.059	0.169	0.638	0.061
JPN (2000)	0.165	0.135	0.030	0.181	0.027	0.277	0.954	0.013
NLD (2000)	0.149	0.059	0.090	0.603	0.078	0.492	0.669	0.163
N.ZLD (2000)	0.183	0.095	0.088	0.482	0.073	0.554	0.357	0.356
NOR (2000)	0.145	0.060	0.085	0.587	0.068	0.438	0.620	0.166
POR (2000)	0.157	0.096	0.061	0.390	0.057	0.212	0.510	0.104
SWE (2000)	0.162	0.051	0.110	0.682	0.077	0.412	0.656	0.142
UK (2000)	0.199	0.087	0.112	0.564	0.065	0.563	0.222	0.438
USA (2000)	0.180	0.137	0.043	0.240	0.026	0.403	0.725	0.111
BEL (mid-1990s)	0.265	0.075	0.190	0.719	0.109	0.360	0.190	0.291
SWI (2000)	0.115	0.058	0.057	0.494	0.065	0.335	1.629	-0.210

Notes: (m) missing data; (1), (2), (3), and (6) - Forster, M. and d'Ercole, M.M. (2005); (5) OECD Social Expenditure Database; (7) transfers and taxes received (paid) by the three bottom deciles of the population as a share of their equivalized disposable income - data kindly provided by Marco Mira D'Ércole based on the OECD questionnaire on distribution of household incomes.

Table 8: Maximum likelihood estimation results (Z1: net targeting)

Production function		Technical inefficiency effects		Variance parameters	
Parameter	Estimate	Parameter	Estimate	Parameter	Estimate
β_0	-5.078	δ_0	-0.341	σ^2	0.037
	(-2.364)		(-0.479)		(0.821)
β_1	-2.310	δ_1	1.101	γ	0.835
	(-2.513)		(0.918)		(3.348)
β_2	-17.283	Log likelihood	24.513		
	(-1.987)				
β_3	-0.391				
	(-4.079)				
β_4	-19.675				
	(-1.990)				
β_5	-2.392				
	(-1.275)				

Table 9: Likelihood ratio tests (H_{01}, H_{02}: 5% level; H_{03}: 10% level)

Null hypothesis	Λ	Critical Value	Decision
H_{01}: $\delta_1=0$	3.648	3.84	Not Reject
H_{02}: $\beta_3=0$; $\beta_4=0$; $\beta_5=0$	18.468	7.82	Reject
H_{03}: $\lambda=0$, $\delta_0=0$, $\delta_1=0$	4.789	5.528	Not Reject

References

Achian, A., Demsetz, H. (1972), "Production, Information Costs, and Economic Organisation", *American Economic Review*, 62, p. 777-795.

Afonso, A. and Fernandes, S. (2006), "Local Government Spending Efficiency: DEA Evidence for the Lisbon Region", *Regional Studies*, 40(1), p. 39-53.

Afonso, A. and St. Aubyn, M. (2004), "Non-parametric Approaches to Public Education and Health Expenditure Efficiency in OECD countries", ISEG/UTL Technical University of Lisbon, Economics Department, Working Paper 1/2004/DE/CISEP/UECE.

Afonso, A. and St. Aubyn, M (2005), "Cross-country Efficiency of Secondary Education Provision: a Semi-parametric Analysis with Non-discretionary Inputs", European Central Bank, Working Paper n. 494.

Afonso, A., Ebert, W., Schuknecht, L., and Thöne, M. (2005), "Quality of public Finances and Growth", European Central Bank, Working Paper Series no. 438.

Afonso, A., Tanzi, V. and Schuknecht, L. (2005), "Public Sector Efficiency: an International Comparison," *Public Choice*, 123 (3-4), p. 321-347.

Agranoff, R. and McGuire, M. (2003), *Collaborative Public Management: New Strategies for Local Governments*, Washington DC, Georgetown University Press.

Aigner, D.J., Chu S.F. (1968), "On Estimating the Industry Production Function", *American Economic Review*, 58, p. 826-839.

Aigner, D.J., Lovell, C.A.K., Schmidt P. (1977), "Formulation and Estimation of Stochastic frontier Production Function Models", *Journal of Econometrics*, 6, p. 21-37.

Alchian, A. and Woodward, S. (1987), "Reflections on the Theory of the Firm", *Journal of Institutional and Theoretical Economics*. 143(1), p. 110–137.

Alchian, A., and Demsetz, H. (1972), "Production, Information Costs, and Economic Organisation"*, American Economic Review*, 62, p. 777-795.

Allais, M. (1943), *Traité d'Economie Pure*, 1952 edition of *A la Recherche d'une Discipline Economique*, Paris, Impremerie Internationale.

Alston, L., Libecap, G. and Mueller, B. (1997), "Violence and the Development of Property Rights to Land in the Brazilian Amazon", in Drobak, J. and Nye, J. (eds), *The Frontiers of the New Institutional Economics*, San Diego, Academic Press, p. 145-164.

Aoki, M. (1986), "Horizontal versusVertical Information Structure of the Firm", *American Economic Review*, 76, p. 971-983.

Aoki, M. (1988), *Information, Incentives, and Bargaining in the Japanese Economy*, Cambridge, Cambridge University Press.

Aoki, M. (1990), "Toward an Economic Model of the Japanese Firm", *Journal of Economic Literature*, 28(1), p. 1-27.

Aoki, M. (2001), *Toward A Comparative Institutional Analysis*, Cambridge, MIT Press.

Araujo, L., Dubois, A. and Gadde, L-E (2003), "The Multiple Boundaries of the Firm"*, Journal of Management Studies*, 40(5), p. 1255-1277.

Arrow, K. (1963), "Uncertainty and the welfare economics of medical care", *American Economic Review*, 53(5), p. 941-973.

Arrow, K. (1964), "Control in Large Organisations", *Management Science*, 10(3), p.397-408.

Arrow, K. (1974), *The Limits of Organisation,* New York, Norton.

Arrow, K. (2001), "Reflections on Reflections", *Journal of Health Politics, Policy and Law*, p. 1197-1204.

Arrow, K. and Debreu, G. (1954), "Existence of an Equilibrium for a Competitive Economy", *Econometrica*, 22, p. 265-290.

Arrow, K. and Lind, R.C. (1970), *Uncertainty and the Evaluation of Public Investment Decisions,* American Economic Review, 60(3), p. 374.

Aschauer, D. (1989), "Is public expenditure productive?", *Journal of Monetary Economics*, 23, p. 177-200.

Atkinson, A.B., Rainwater, L. and Smeeding, T. (1995), "Income Distribution in OECD Countries", OECD, Paris.

Atkinson, T. (2005), "Atkinson Review (final report) – measurement of government output and productivity for the national accounts", London, Palgrave.

Bajari, P and Tadelis, S. (2001), "Incentives versus Transaction Costs: A Theory of Procurement Contracts", *Rand Journal of Economics*, 32(3), p. 387-407.

Baker, G., Gibbons, R. and Murphy, K. (2002), "Relational Contracts and the Theory of the Firm", *Quarterly Journal of Economics*, 117(1), p. 39-84.

Baker, G., Gibbons, R. and Murphy, K. (2003), "Relational Contracts in Strategic Alliances", unpublished manuscript, Cambridge, MIT, Sloan School of Management.

Baldwin, P. (1990), *The Politics of Social Solidarity: Class Bases of the European Welfare State 1875-1975*, Cambridge University Press, Cambridge.

Baldwin, R. And Cave, M. (1999), *Understanding Regulation: Theory, Strategy, and Practice,* Oxford, Oxford University Press.

Banker, R.D., Charnes, A., and Cooper, W.W. (1984), "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis", *Management Science*, 30, p. 1078-1092.

Barnard, C.I. (1938), *The Functions of the Executive*. Cambridge (MA), Harvard University Press.

Barney, J. (1999), "How a firm's capabilities affect boundary decisions", *Sloan Management Review,* 40(3), p. 137-47.

Barone, E. (1908), "The Ministry of Production in the Collectivist State", *Giornale degli Economisti*, as translated in Hayek (1935), ed., *Collectivist Economic Planning*, London, Routledge.

Barr, N. (1992), "Economic Theory and the Welfare State: A Survey and Interpretation", Welfare State Discussion Paper 54, STICERD, London School of Economics, London.

Barro, R. and Lee, J-W. (2001), "Schooling Quality in a Cross-Section of Countries", *Economica*¸ 68 (272), p. 465-488.

Barro, R.J and Sala-I-Martin, X. (1995), *Economic Growth*, New York, McGraw-Hill.

Barro, R.J. (1990), "Government spending in a simple model of endogenous growth", *Journal of Political Economy*, 98(5), p. 103-125.

Barro, R.J. (1991), "Economic growth in a cross section of countries", *Quarterly Journal of Economics*, 106(2), p. 407-443.

Barro, R.J. (2001), Education and economic growth. In Helliwell, J.F. ed., *The Contribution of Human and Social Capital to Sustained Economic Growth and Well-Being*, OECD, chapter 3, pp. 14-41.

Barth, J. and Bradley, M. (1988), *The impact of government spending on economic activity*, The National Chamber Foundation, Washington D.C.

Bassanini, A. and Scarpetta, S. (2001), "Does human capital matter for growth in OECD countries? Evidence from pooled mean-group estimates", OECD Economics Department Working Paper 282, Paris.

Bassanini, A.; Scarpetta, S. and Hemmings, P. (2001), "Economic growth: the role of policies and institutions. Panel data evidence from OECD countries", OECD Economics Department Working Paper 283, Paris.

Bastiat, F. (1944-45), *Harmonies of Political Economy,* 2 vs. Register Publishing Co., Santa Ana, California.

Battese, G.E. and Coelli, T.J. (1995), "A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data", *Empirical Economics*, 20, p. 325-332.

Battese, G.E. and Corra, G.S. (1977), "Estimation of a Production Frontier Model: With Application to the Pastoral Zone of Eastern Australia", *Australian Journal of Agricultural Economics*, 21, p. 169-179.

Becker, G.S. (1975), *Human Capital: a theoretical and empirical analysis, with special reference to education,* New York, National Bureau of Economic Research, Columbia University Press.

Becker, G.S. (1993), *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*, 3rd edition, Chicago, University of Chicago Press (1st edition published in 1964).

Beckerman, W. (1979), "The Impact of Income Maintenance Payments on Poverty inBritain – 1975", *Economic Journal*, June, p. 261-279.

Beckerman, W. and Clark (1982), *Poverty and Social Security in Britain since 1961*, Oxford University Press, London.

Beckmann, M.J. (1988), *Tinbergen Lectures on Organisational Theory*, New York, Springer Verlag.

Behrman, J.R., Deolalikar, A.B. and Soon, L-Y (2002), "Conceptual Issues in the Role of Education Decentralization in Promoting Effective Schooling in Asian Developing Countries", ERD Working Paper Series No. 22, Asian Development Bank.

Benhabib, J. and Spiegel, M.M. (1994), "The role of human capital in economic development: evidence from aggregate cross-country data", *Journal of Monetary Economics*, 34(2), p. 143-174.

Benz, A. (2004), "Einleitung: Governance - Modebegriff oder nützliches sozialwissenschaftliches Konzept?" in *Governance - Regieren in komplexen Regelsystemen: Eine Einführung*, Benz, A. (ed.), Verlag für Sozialwissenschaften.

Besley, T. and A. Case. (1995). "Incumbent Behavior: Vote-Seeking, Tax-Setting, and Yardstick Competition," American Economic Review 85, 25–45.

Bils, M. and Klenow, P.J. (2000), "Does schooling cause growth?", *American Economic Review*, 90(5), p. 1160-1183.

Bleaney, M.; Gemmell, N. and Kneller, R. (2001), "Testing the Endogenous Growth Model: Public Expenditure, Taxation and Growth over the Long Run", *Canadian Journal of Economics*, 34 (1), p. 36-57.

Bloom, D.; Canning, D. and Sevilla, J. (2001), "The Effect of Health on Economic Growth: Theory and Evidence", NBER Working Paper No. 8587, Cambridge MA.

Bolton, P. and Farrel, J. (1990), "Decentralization, Duplication, and Delay", *Journal of Political Economy*, 98, August, p. 803-826.

Bond, S. (2002), "Dynamic panel data models: a guide to micro data methods and practice", *Portuguese Economic Journal*, 1, p. 141-162.

Bordignon, M., F. Cerniglia and F. Revelli. (2002). "In Search for Yardstick Competition: Property Tax Rates and Electoral Behavior in Italian Cities," Munich: Cesifo Working Paper No. 644(1).

Boyne, G. (1998), "Public Services Under New Labour: Back to Bureaucracy?", *Public Money & Management*, July-September, p. 43-50.

Boyne, G., Farrell, C., Law, J., Powell, M. and Walker, R.M. (2003), *Evaluating Public Management Reforms*, Buckingham, Open University Press.

Bradach, J.L. and Eccles, R. (1989), "Price, authority and trust: from ideal types to plural forms", *Annual Review of Sociology*, 15, p. 97-118.

Bradshaw, J., Ditch, J., Holmes, H. and Whiteford, P. (1993), "Support for Children: A Comparison of Arrangements in Fifteen Countries", Report to the Department of Social Security, University of York, March.

Buchanan, J.M. (1985), *The moral dimension of debt financing*, Economic Inquiry, 13, p. 1-6.

Buchanan, J.M. and Stubblebine, W.C. (1962), "Externality", *Economica*, 29(116), p.371-384.

Buck, D, (2000), "The efficiency of the Community Dental Service in England: a, data envelopment analysis", *Community Dentistry and Oral Epidemiology*, 28 (4), p. 274-280.

Burau, V. and Blank, R.H (2006), "Comparaing Health Policy: An Assessment of Typologies of Health Systems", *Journal of Comparative Policy Analysis*, 8(1), p. 63-76.

Busom, I. (1999), "An Empirical Evaluation of the Effects of R&D Subsidies", Working Paper No. B99-05, Universitat Autonoma de Barcelona.

Buysse, K. (2002), "Human Capital and Growth in OECD Countries: the Role of Public Expenditure on Education", mimeo, paper presented at the 4th Public Finances Workshop of the Banca d'Italia.

Caiden, G. (1988), "The vitality of administrative reform", *International Review of Administrative Sciences*, 7, p. 331-57.

Cameron, S.V. and Heckman, J.J, (1999), "Can tuition policy combat rising wage inequality?", in Kosters, M.H. (ed), *Financing College Tuition: Government Policies and Educational Priorities*, Washington, AEI Press.

Carneiro, P. and Heckman, J.J. (2002), "The Evidence on Credit Constraints in Post-Secondary School", Discussion paper no. 518, the Institute for the Study of Labor (IZA), Germany.

Cashin, P. (1995), "Government spending, taxes and economic growth", IMF Staff Papers, 42, p. 237-269.

Cassou, S. and Lansing, K. (1999), "Fiscal Policy and Productivity Growth in the OECD", *Canadian Journal of Economics*, 32, p. 1215-1226.

Castles, F.G. and Mitchell, D. (1993), "Worlds of Welfare and Families of Nations", in Castles, F.G. (ed.), *Families of Nations*, Dartmouth, Aldershot, p. 93-128.

Chandler, A.D. Jr. (1977), *The Visible Hand: The Managerial Revolution in American Business,* Cambridge, MA, The Belknap Press.

Charnes, A., Cooper, W.W. and Rhodes, E. (1978), "Measuring the Efficiency of Decision Making Units", *European Journal of Operational Research*, 2, p. 429-444.

Cheung, S. (1983), "The Contractual Nature of the Firm", *Journal of Law and Economics*, 26(1), p. 1-22.

Clements, B. (1999), "The Efficiency of Education Expenditure in Portugal", IMF working paper nº 99/179.

Clements, B. (2002), "How Efficient is Education Spending in Europe?" *European Review of Economics and Finance*, 1 (1), p. 3-26.

Coase, R.H. (1937), "The Nature of the Firm", *Economica*, 4, November, p. 36-405.

Coase, R.H. (2000), "The Acquisition of Fisher Body by General Motors", *Journal of Law and Economics*, 43(1), p. 15-32.

Coase, R.H. (2005), "The new institutional structure of production", chapter 2 in Menard, C. and Shirley, M.M. (eds.), *Handbook of New Institutional Economics*, Springer, p. 31-39.

Coelli, T. (1996a). "Assessing the Performance of Australian Universities using Data Envelopment Analysis," *mimeo*, University of New England.

Coelli, T. (1996b), "A Guide to FRONTIER Version 4.1: A Computer Program for Stochastic Frontier Production and Cost Function Estimation", CEPA Working Paper 96/07.

Coelli, T. and Perelman, S. (2000), "Technical efficiency of European railways: a distance function approach", *Applied Economics*, 32, p. 1967-1976.

Coelli, T., Rao, D.S.P. and Battese, G.E. (1998), *An Introduction to Efficiency and Productivity Analysis*, London, Kluwer Academic Publishers.

Cohn, E. and Geske, T.G. (1990), *The Economics of Education*, 3rd edition, Oxford, Pergamon Press.

Commons, J. (1934), *Institutional Economics. Its Place in Political Economy*, New York, Macmillan.

Cook, M.L. (1995), "The Future of US Agricultural Cooperatives: A Neo-institutional Approach", *American Journal of Agricultural Economics*, 77, December, p. 1153-1159.

Cook, M.L. and Iliopoulos, C. (2000), "Ill-defined Property Rights in Collective Action: The Case of US Agricultural Cooperatives", in Menard, C. (ed.), *Institutions, Contracts and Organisations. Perspectives from New Institutional Economics*, Cheltenham, E. Elgar, p. 335-348.

Cooper, W.W., Seiford, L.M. and Tone, K., (2000), *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*, Boston, Kluwer Academic Publishers.

Culter, D.M. (2002), "Health Care and the Public Sector", in Auerbach, A.J. and Feldstein, M. (eds), *Handbook of Public Economics*, 4, Elsevier, p. 2143-2243.

Cutler, D.M. and Horwitz, J. (1999), Converting hospitals from not-for-profit to for-profit form: why and what effects?", in Cutler, D. (ed.), *The Changing Hospital Industry: Comparing Not-for-Profit and for-Profit Institutions*, Chicago, University of Chicago Press, p. 45-79.

Dar, A.A. and AmirKhalkhali, S. (2002), "Government size, factor accumulation, and economic growth: evidence from OECD countries", *Journal of Policy Modelling* (24/7-8), p. 679-692.

Darby, M.R. and Karni, E. (1973), "Free competition and the optimal amount of fraud", *Journal of Law & Economics*, 16, p. 67-88.

Dawson, D., Gravelle, H., Kind, P., O'Mahony, M., Street, A., Weale, M. (2004), "Developing New Approaches to Measuring NHS Outputs and Productivity – First Interim Report to Department of Health", CHE Technical Paper Series 31.

De Borger, B. and Kerstens, K. (1996). "Cost Efficiency of Belgian Local Governments: A Comparative Analysis of FDH, DEA, and Econometric Approaches," *Regional Science and Urban Economics*, 26(2), p. 145-170.

De Gregorio, J. (1996), "Inflation, Growth and Central Banks: Theory and Evidence", The World Bank, Working Paper 1575, Washington D.C.

De la Fuente, A. (1997), "Fiscal Policy and Growth in the OECD", CEPR Discussion, DP1755.

De la Fuente, A. and Domenech, R. (2000), "Human Capital in Growth Regressions: How Much Difference Does Data Quality Make?", OECD Economics Department Working Paper No. 262.

Deaton, A. and Muellbauer, J. (1980), *Economics and Consumer Behaviour*, Cambridge University Press.

Debreu, G. (1951), "The Coefficient of Resource Utilisation", *Econometrica*, 19, p. 273-292.

Demsetz, H. (1988a), "The Theory of the Firm Revisited", *Journal of Law, Economics, and Organisation.* 4(1), p. 141-161.

Demsetz, H. (1988b), *The Organisation of Economic Activity*, 2 vols., Oxford, Basil Blackwell.

Demsetz, H. (1988c), *Ownership, control and the firm*, Basil Blackwell, Oxford.

Demsetz, H. (1995) *The Economics of the Business Firm. Seven Critical Commentaries*, Cambridge, Cambridge University Press.

Demsetz, H. (2002), "Theories of the Firm and Externalities: A Critical Evaluation of the Role of Transaction Costs". WP. 20 pages.

Deprins, D., Simar, L., and Tulkens, H. (1984), "Measuring labor-efficiency in post offices", in Marchand, M. Pestieau, P. and Tulkens, H. (eds.), *The performance of public enterprises: concepts and measurement*, Amsterdam, North-Holland.

Devarajan, S.; Swaroop, V. and Zou, Z. (1996), "The composition of public expenditure and economic growth", *Journal of Monetary Economics*, 37, p. 313-344.

Diamond, A. (1999), "Does Federal Funding "Crowd In" Private Funding of Science?", *Contemporary Economic Policy*, 17, p. 423-431.

Diedrichs, U. (2004), "Overview Paper on Classification and Mapping of Governing Modes", NEWGOV, project 1: The Evolution and Impact of Governing Modes, reference number 1/D08.

Dixon, J. (1999), *Social Security in Global Perspective*, London, Westport.

Docteur, E., Oxley, H. (2003), "Health-Care Systems: Lessons from the Reform Experience", Health Working Papers No. 9., OECD, Paris.

Dor, A. (1994), "Non-minimum cost functions and the stochastic frontier: on applications to health care providers", *Journal of Health Economics*, 13(3), p. 329-34.

Dow, G. (1987), "The Function of Authority in Transaction Costs Economics", *Journal of Economic Behavior and Organisation*, 8(1), p. 13-38.

Doz, Y.L. and Hamel, G. (1998), *Alliance Advantage: The Art of Creating Value through Partnering,* Boston, Massachusetts, Harvard Business School Press.

Dunleavy, P. and Hood, C. (1994), "From old public administration to new public management", *Public Money and Management*, 14(3), p. 9-16.

Dupuit, J. (1844), "De la Measure de l'Utilité des Travaux Publiques", *Annales des Ponts et Chaussées 8*, translated and reprinted in K. Arrow and T. Scitovski (eds.), 1969, AEA Readings in Welfare Economics, AEA, p. 255-283.

Dwyer, F.R. and Oh, S. (1988), "A Transaction Cost Perspective on Vertical Contractual Structure and Interchannel Competitive Strategies", *Journal of Marketing*, 52, p. 21-34.

Dyer, J.H. (1997), "Effective Interfirm Collaboration: How Firms Minimize Transaction Costs and Maximize Transaction Value", *Strategic Management Journal*, 18(7), p. 535-556.

Easterly, W. and Rebelo, S. (1993), "Fiscal Policy and Economic Growth: An Empirical Investigation", *Journal of Monetary Economics*, 32(3), p. 417-458.

Eccles, R. (1981), "The Quasifirm in the Construction Industry", *Journal of Economic Behavior and Organisation*, 2(4), p. 335-357.

Eccles, R.G. and Crane, D. (1987), "Managing through networks in investment banking", *California Management Review*, 30(1), p. 176-195.

Ehrlich, I. (1975), On the relation between education and crime", in Juster, F.T. (ed), *Education, Income, and Human Behavior*, New York, McGraw-Hill.

Englander, S. and Gurney, A. (1994), "Medium-term determinants of OECD Productivity", OECD Economic Studies, No. 22, Paris.

Ercole, M. and Salvini, A. (2003), "Towards Sustainable Development: the Role of Social Protection", OECD Social, Employment and Migration Working Papers, nº 12.
Esping-Andersen, G. (1990), *The three worlds of welfare capitalism,* Polity Press, Cambridge.

European Commission (2004), "European Competitiveness Report", Commission Staff Working Document, SEC(2004)1397.

Evans, D., Tandon, A., Murray, C. and Lauer, J. (2000). "The Comparative Efficiency of National Health Systems in Producing Health: an Analysis of 191 Countries," GPE Discussion Paper Series 29, World Health Organisation, Geneva.

Evans, D., Tandon, A., Murray, C. and Lauer, J. (2003), Health System Efficiency: Concepts, chapter 50 in Evans D.B. and Murray, C.J.L. (eds), *Health Systems*

Performance Assessment: Debates, Methods and Empiricism, World Health Organisation.

Fakin, B. and de Crombrugghe, A. (1997). "Fiscal adjustment in transition economies: social transfers and the efficiency of public spending: a comparison with OECD countries," Policy Research Working Paper 1803. Washington, DC: World Bank.

Färe, R. and Grosskopf, S. (1996), *Intertemporal Production Frontiers: with Dynamic DEA*, Dordrecht, Kluwer.

Färe, R. and Lovell, C.A.K. (1978), "Measuring the Technical Efficiency of Production", *Journal of Economic Theory*, 19, p. 150-162.

Farrell, J. and Scotchmer, S. (1988), "Partnerships", *Quarterly Journal of Economics*, 103, May, p. 279-297.

Farrell, M. (1957). "The Measurement of Productive Efficiency," *Journal of the Royal Statistical Society*, Series A, 120, Part 3, p. 253-290.

Farsi, M., Filippini, M. and Kuenzle, M. (2003), *Unobserved Heterogeneity in Stochastic Cost Frontier Models: a Comparative Analysis*, Oviedo, University of Oviedo.

Faulkner, D.O. and De Rond, M. (2000), *Cooperative Strategy: Economic, Business and Organisational Issues,* Oxford, Oxford University Press.

Ferrera, M. (1996), "The Southern Model of Welfare in Social Europe", *Journal of European Social Policy*, 6(1), p. 17-37.

Ferrera, M. (1998), "The Four "Social Europes": Between Universalism and Selectivity", in Mény, Y. and Rhodes, M. (eds.), *The Future of Welfare in Europe: A New Social Contract?,* Palgrave, Houndmills, p. 79-96.

Fölster, S., Henrekson, M. (1999), "Growth and the Public Sector: A Critique of the Critics", *European Journal of Political Economy*, 15, p. 337-358.

Ford, R. and Poret, P. (1991), "Infrastructure and private-sector productivity", OECD Economics Department Working Paper 91, Paris.

Forster, M. (1993), "Comparing Poverty in 13 OECD Countries – Traditional and Synthetic Approaches", Luxembourg Income Study, Working Paper series, 100, Syracuse University, New York.

Forster, M. and d'Ercole, M.M. (2005), "Income Distribution and Poverty in OECD Countries in the second half of the 1990s", Social, Employment and Migration Working Paper No. 22, OECD, Paris.

Förster, M. and Pearson, M. (2002), "Income Distribution and Poverty in the OECD Area: Trends and Driving Forces", OECD Economic Studies No. 34.

Foss, N.J., Foss, K., Klein, P.G. and Klein, S.K. (2002), "Heterogeneous Capital, Entrepreneurship, and Economic Organisation", *Journal des Economistes et des Etudes* Humaines, 12(1), p. 79-96.

Fried, H.O., Lovell, C.A.K., Schmidt, S.S., Yaisawarng, S. (2002), "Accounting for environmental effects and statistical noise in, data envelopment analysis", *Journal Of Productivity Analysis*, 17(1-2), p. 157-174.

Gemmell, N., Kneller, R. (2001), "The impact of fiscal policy on long-run growth", *European Economy*, 1, p. 97-129.

Gerdtham, U.G., Rehnberg, C., and Tambour, M. (1999), "The impact of internal markets on health care efficiency: Evidence from health care reforms in Sweden", Applied Economics, 31, p. 935-945.

Goldberg,V.P. (1980), "Relational Exchange: Economics and Complex Contracts", *American Behavioral Scientist*, 23(3), p. 337-352.

Gong, B.H., and Sickles, R.C. (1992), "Finite Sample Evidence on the Performance of Stochastic Frontier Models Using Panel Data," *Journal of Productivity Analysis*, 1, p. 229-61.

Gouyette, C. and Pestieu, P. (1999), "Efficiency of the Welfare State", Kyklos, 52, p. 537-553.

Grandori, A. and Soda, G. (1995), "Inter-firm Networks: Antecedents, Mechanisms and Forms", *Organisation Studies*, 16(2), p. 183-214.

Greene, W. (2004), "Distinguishing between heterogeneity and inefficiency: stochastic frontier analysis of the World Health Organisation's panel data on national health care systems", Health Economics, 13(10), p. 959-980.

Greene, W. H. (2005), "Reconsidering Heterogeneity in Panel Data Estimators of the Stochastic Frontier Model", *Journal of Econometrics*, 126(2), p. 269-303.

Grosse, T.G. (2005), *Inception Report: Democratization, Capture of the State and New Forms of Governance in CEE Countries*, NEWGOV Deliverables, Reference Number 17/D2.

Grossman, S. and Hart, O. (1986). "The costs and benefits of integration: a theory of vertical and lateral integration", *Journal of Political Economy*, 94 (4), p. 691-719.

Guellec, D. and van Pottelsberghe de la Potterie, B. (2000), "The Impact of Public R&D Expenditure on Business R&D", OECD STI Working Paper 2000/4, Paris.

Gupta, S. and Verhoeven, M. (2001), "The Efficiency of Government Expenditure – Experiences from Africa", *Journal of Policy Modelling*, 23, p. 433-467.

Gwartney, J., Lawson, R. and Holcombe, R. (1998), "The size and functions of government and economic growth", study prepared for the Joint Economic Committee of the US Congress, Washington, DC, April.

Hakansson, H., and Johanson, J. (1993), "The Network as a Governance Structure: Interfirm Cooperation Beyond Markets and Hierarchies", in Grabher, G. (ed.), *The Embedded Firm: On the Socioeconomics of Networks*, London, Routledge, p. 35-51.

Hansmann, H. (1980), "The role of nonprofit enterprise", *Yale Law Journal*, 89, p. 835-901.

Hansmann, H. (1988), "The Ownership of the Firm", *Journal of Law, Economics and Organisation*, 4(2), p. 267-304.

Hanushek, E. (2002), "Publicly Provided Education", in Auerbach, A.J. and Feldstein, M. (eds.), *Handbook of Public Economics*, 4, North-Holland.

Hanushek, E. and Kimko, D. (2000), "Schooling, labor force quality, and economic Growth", *American Economic Review*, 90 (5), p. 1184-1208.

Hanushek, E. and Luque, J. (2003), "Efficiency and equity in schools around the World", *Economics of Education Review*, 22, p. 481-502.

Hanushek, E. and Taylor, L. (1990), "Alternative Assessments of the Performance of Schools", *The Journal of Human Resources*, 25(2), p. 179, 201.

Hart, O.D. and Moore, J. (1990), "Property Rights and the Nature of the Firm", *Journal of Political Economy*, 98, p. 1119-1158.

Haveman, R.H., and Wolfe, B.L. (1984), "Schooling and economic well-being: the role of nonmarket effects", *Journal of Human Resources*, 19(3), p. 377-407.

Hayek, F. (1945), "The Use of Knowledge in Society", *American Economic Review*, 35 (4), p. 519–530.

Hayek, F.A. (1973), *Law, Legislation and Liberty: A new statement of the liberal principles of justice and political economy,* Volume 1, *Rules and Order*, London and Henley, Routledge & Kegan Paul.

Head, J.G. (1974), *Public Goods and Public Welfare,* Duke University Press, Durham, North Carolina.

Heitger, B. (2001), "The Scope of Government and Its Impact on Economic Growth in OECD Countries", Kiel Working Paper No. 103, Kiel.

Hemming, R., Kell, M. and Mahfouz, S. (2002), "The Effectiveness of Fiscal Policy in Stimulating Economic Activity - A Review of the Literature", IMF Working Paper No. 02/208.

Henderson, D.J. (2003), "The Measurement of Technical Efficiency Using Panel Data", Working Paper No. 0308, Binghamton University, Department of Economics.

Heritier, A. (2002), "New Modes of Governance in Europe: Policy-Making without Legislating?", in Heritier, A. (ed.), *Common Goods: Reinventing European and International Governance*, Lanham, Rowman and Littlefield.

Hill, M. (1996), *Social Policy: a comparative analysis*, Prentice-Hall.

Hollingsworth, B. and Wildman, J. (2002), "Efficiency and Cross Efficiency Measures: A Validation Using OECD Data", Centre for Health Program Evaluation, Working Paper No. 132.

Holmström, B. and Roberts, J. (1998), "The Boundaries of the Firm Revisited", *Journal of Economic Perspectives*, 12 (4), p. 73-94.

Holtz-Eakin, D. (1994), "Public sector capital and the productivity puzzle", *Review of Economics and Statistics*, 76, p. 12-21.

Hotelling, H. (1938), "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates", *Econometrica*, 6, p. 242-69.

Ioannides, S. (2003), "Orders and Organisations: Hayeian Insghts for a Theory of Economic Organisation", *American Journal of Economics and Sociology*, 62(3), p. 533-565.

Jachtenfuchs, M. and Kohler-Koch, B. (2004), "Governance and Institutional Development", in Wiener, A. and Dietz, T. (eds), *European Integration Theory*, p. 97-116, Oxford, Oxford University Press.

Jacobs, R. (2001), "Alternative methods to examine hospital efficiency: data envelopment analysis and stochastic frontier analysis", *Health Care Management Science*, 2, p. 103-115.

Jacobs, R., Smith, P.C. and Goddard, M. (2003), "Measuring performance: An examination of composite performance indicators", A report for the Department of Health, Centre for Health Economics, University of York.

Jessop, B. (1998), "The rise of governance and the risks of failure: the case of economic development", *International Social Science Journal*, 50(155), p. 29-45.

Johnes, G., Bradley S., and Millington, J. (2001), "The effect of competition on the efficiency of secondary schools in England", European Journal of Operational Research, 135, p. 545-568.

Johnes, G., Izadi, H., Crouchley, R. and Ozcrochi, R. (2002), "Stochastic frontier estimation of a CES cost function: the case of higher education in Britain", Economics of Education Review, 21, p. 63-72.

Johnston, R. and Lawrence, P. (1988), "Beyond vertical integration – the rise of the value-adding partnership", *Harvard Business Review*, 88, p. 94-104.

Jorde T. and Teece, D. (1989), "Competition and Cooperation: Striking the Right Balance", *California Management Review*, 31(3), p. 25-37.

Jorgenson, D.W. and Fraumeni, B.M. (1992), "Investment in education and U.S. economic growth", *Scandinavian Journal of Economics*, 94, p. 51-70.

Joskow, P. (1985), "Vertical Integration and Long-Term Contracts: The Case of Coal-Burning Electric Generating Plants", *Journal of Law, Economics, and Organisation*, 1(1), p. 33-80.

Joskow, P. (2005), "Vertical Integration", chapter 13 in Menard, C. and Shirley, M.M. (eds.), *Handbook of New Institutional Economics*, Springer.

Joumard, I. (2002), "Tax Systems in the European Union Countries", OECD Economic Studies, 34, p. 91-151.

Juhn, C., Murphy, K.M., and Pierce, B. (1993), "Wage inequality and the rise in returns to skill", *Journal of Political Economy*, 101(3), p. 410-442.

Kalyvitis, S., and Kalaitzidakis, P. (2002), "Financing 'New' Public Investment and/or Maintenance in Public Capital for Long-run Growth? The Canadian Experience", Athens University of Economics and Business, Athens.

Kamps, C. (2004), "New Estimates of Government Net Capital Stocks for 22 OECD Countries 1960-2001", IMF Working Paper 04/67, International Monetary Fund. Washington D.C.

Keefer, P., Knack, S. (1997), "Why don't poor countries catch up? A cross-country test of an institutional explanation", *Economic Inquiry*, 35, p. 590-602.

Kenworthy, L. (1998), "Do Social-Welfare Policies Reduce Poverty? A Coss-National Assessment", Luxembourg Income Study, Working Paper Series, 35.

Kersbergen, K. and Waarden, F. (2004), "Governance as a Bridge Between Disciplines: Cross-disciplinary Inspiration Regarding Shifts in Governance and Problems of Governability, Accountability and Legitimacy", *European Journal of Political Research*, 43, p. 143-171.

Kickert, W. (ed.) (1997), *Public Management and Administrative Reform in Western Europe*, Cheltenham, Edward Elgar.

Kickert, W.J.M., Klijn, E.-H. and Koppenjan, J.F.M. eds. (1997), *Managing Complex Networks*, London, Sage Publications.

Killing, J.P. (1988), "Understanding Alliances: The Role of Task and Organisational Complexity", in Contractor, F.J. and Lorange, P. (eds.), *Cooperative strategies in International Business,* New York, Lexington Books, p. 55-67.

Klein, B., Crawford, R.G. and Alchian, A.A. (1978), "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process", *Journal of Law and Economics*, 21, p. 297-326.

Kneller, R.; Bleaney, M. and Gemmell, N. (1998), "Growth, Public Policy and the Government Budget Constraint: Evidence from OECD Countries", School of Economics, University of Nottingham, Discussion Paper No. 98/14, Nottingham.

Kodde, D.A. and Palm, F.C. (1986), "Wald criteria for jointly testing equality and inequality restrictions", *Econometrica*, 54(5), p. 1243-1248.

Kogut, B. (2000), "The network as knowledge: generative rules and the emergence of structure", *Strategic Management Journal*, 21(3), p. 405-25.

Kohler-Koch, B. (1999), "The Evolution and Transformation of European Governance", in Kohler-Koch, B. and Eising, R. (eds.), *The Transformation of Governance in the European Union*, London, Routledge.

Komai, J. (1986), "The Soft Budget Constraint", *Kyklos*, 39(1), p. 3-30.

Kooiman, J. (2003), *Governing as Governance*, London, Sage.

Koopmans, T.C. (1951), "An Analysis of Production as an Efficient Combination of Activities", in Koopmans, T.C. (ed.) *Activity Analysis of Production and Allocation*, Cowles Commission for Research in Economics, Monograph no. 13, New York, Wiley.

Kooreman, P. (1994), "Nursing home care in The Netherlands: A nonparametric efficiency analysis", *Journal of Health Economics*, 13, p. 301-16.

Korpi, W. (1983), *The Democratic Class Struggle,* Routledge and Kegan Paul, London.

Korpi, W. and Palme, J. (1998), "The Paradox of Redistribution and Strategies of Equality: Welfare State Institutions, Inequality, and Poverty in the Western Countries", *American Sociological Review*, 63, p. 661-687.

Kraus, M. (2004), "Social Security Strategies and Redistributive Effects in European Social Transfer Systems", *Review of Income and Wealth*, 50(3), p. 431-457.

Kumbhakar, S.C., Ghosh, S. and McGukin, J.T. (1991), "A Generalised Production Frontier Approach for Estimating Determinants of Inefficiency in U.S. Dairy Farms", *Journal of Business and Economics Statistics*, 9, p. 279-286.

Kumbhaker, S.C. and Lovell, C.A.K. (2000), *Stochastic Frontier Analysis*, Cambridge, Cambridge University Press.

Lafontaine, F. and Slade, M. (1997), "Retail Contracting: Theory and Practice", *Journal of Industrial Economics*, 45, March, p. 1-25.

Lancaster, K. (1971), *Consumer Demand: A New Approach*, New York, Columbia University Press.

Lane, J.E. (1997), *Public Sector Reform*, London, Sage.

Lane, J.E. (2000), *New Public Management*, London, Routledge.

Lange, O. (1942), "The Foundations of Welfare Economics", *Econometrica*, 10 (3), p. 215-28.

Langlois, R.N. (1992), "Orders and Organisations: Toward an Austrian Theory of Social Institutions", in Caldwell and Boehm (eds.), *Austrian Economics: Tensions and New Directions,* Boston, MA, Kluwer.

Lee, J-W. (1995), "Capital goods imports and long-run growth", *Journal of Development Economics*, 48, p. 91-110.

Leemans, A. (ed.) (1976), *The Management of Change in Government*, The Hague, Martinus Nijhoff.

Leibowitz, A. (1974), "Home investments in children", *Journal of Political Economy*, 82(2), p. S111-S131.

Lerner, A.P. (1934), "The Concept of Monopoly and the Measurement of Monopoly Power", *Review of Economic Studies*.

Lewin, L. (1991), *Self Interest and Public Interest in Western Politics*, Oxford, Oxford University Press.

Lewis, J. (1992), "Gender and the Development of Welfare Regimes", *Journal of European Social Policy*, 2(3), p. 159-173.

Llewellyn, K.N. (1931), "What Price Contract? An Essay in Perspective", *Yale Law Review*, 40, p. 704-51.

Loasby, B. J. (1998a). "The organisation of capabilities", *Journal of Economic Behavior & Organisation*, 35(2), p. 139-60.

Loasby, B. J. (1998b). "The concept of capabilities", in Foss, N.J. and Loasby, B.J. (eds.), *Economic Organisation, Capabilities and Co-ordination: Essays in Honour of G. B. Richardson*, London: Routledge, p. 163-82.

Lochner, L. and Moretti, E. (2001), "The effect of education on crime: evidence from prison inmates, arrests, and self-reports", Working Paper 8605, National Bureau of Economic Research, Cambridge, MA.

Lothgren, M. (2000), "Specification and estimation of stochastic multiple-output production and technical inefficiency", *Applied Economics*, 32, p. 1533-1540.

Lucas, R. (1988), "On the mechanisms of economic development", *Journal of Monetary Economics*, 22, p. 3-42.

Lucena, D. and Macedo, J.B. (1996), "Reforming Social Security: Efficiency and Governance", Faculdade de Economia, Universidade Nova de Lisboa, Working Paper nº 270.

Luhmann, N. (1979), *Trust and Power*, New York, Wiley.

Luoma, K., Jarvio, M.L., Suoniemi, I., and Hjerppe, R.T. (1996), "Financial incentives and productive efficiency in Finnish health centres", *Health Economics*, 5(5), p. 435-45.

Mahler, V. and Jesuit, D. (2004), "State Redistribution in Comparative Perspective: A Cross-National Analysis of the Developed Countries", Luxembourg Income Study, Working Paper series, 392.

Mandell, M.P. (2001), *Getting Results through Collaboration: Networks and Network Structures for Public Policy and Management*, Westport, Quorum Books.

Mankiw, N.G., Romer, D., and Weil, D. (1992), "A contribution to the empirics of Economic growth", *Quarterly Journal of Economics*, 107(2), p. 407-437.

Marin, B. and Mayntz, R. (eds.) (1991), *Policy Networks: Empirical Evidence and Theoretical Considerations*, Frankfurt/M., Campus.

Marsh, D. and Rhodes, R.A.W. (1992), *Policy Networks in British Government*, Oxford, Oxford University Press.

Masten, S.E. (1984), "The Organisation of Production: Evidence from the Aerospace Industry", *Journal of Law and Economics*, 27, p. 403-417.

Masten, S.E., Meehan, J. and Snyder, E. (1991), "The Costs of Organisation", *Journal of Law, Economics, and Organisation*. 7(1), p. 1-25.

Mayntz, R. (2003), "From Government to Governance: Political Steering in Modern Societies", paper presented at the International Summer Academy organised by the Institut für ökologische Wirtschaftsforschung gG,bH titled "From Government to Governance: The case of Integrated Product Policy (IPP), Würzburg, September 7-11.

Mayntz, R. (2004), "Governance Theorie als fortentwickelte Steuerungstheorie?", MPIfG Working Paper 04/1, March, Köln, Max-Planck-Institut für Gesellschaftsforschung.

Mayntz, R. and Scharpf, F.W. (1995), "Steuerung und Selbstorganisation in staatsnahen Sektoren", in Mayntz, R. and Scharpf, F.W. (eds.), *Gesellschaftliche Selbstregulierung und Politische Steuerung*, Frankfurt/M., Campus.

Meeusen, W., van den Broeck, J. (1977), "Technical Efficiency and Dimension of the Firm: Some Results on the Use of Frontier Production Functions", *Empirical Economics,* 2(2), p. 109-22.

Ménard, C. (1994), "Organisations as Coordinating Devices", *Metroeconomica*, 45(3), p. 224-247.

Ménard, C. (1996), "On Clusters, Hybrids and other Strange Forms. The Case of the French Poultry Industry", *Journal of Institutional and Theoretical Economics*, 152(1), p. 154-183.

Ménard, C. (1997), "Le Pilotage des Formes Organisationnelles Hybrides", *Revue Economique*, 48(2), p. 741-751, english translation, "The Governance of Hybrid Organisational Forms", in Ménard (ed.) (2004b), volume IV.

Ménard, C. (2004), "The Economics of Hybrid Organisations", *Journal of Institutional and Theoretical Economics*, 160(3), forthcoming.

Ménard, C. (2005), *Modes of Governance*, chapter 12 in Menard, C. and Shirley, M.M. (eds.), *Handbook of New Institutional Economics*, Springer, p. 281-318.

Ménard, C. and Shirley, M.M. (2005), *Introduction*, in Menard, C. and Shirley, M.M. (eds.), *Handbook of New Institutional Economics*, Springer, p. 1-18.

Michael, R.T. (1982), "Measuring non-monetary benefits of education: a survey", in McMahon, W.W. and Geske, T.G. (eds), *Financing Education: Overcoming Inefficiency and Inequity*, University of Illinois Press, Urbana, IL, p. 119-149.

Miles, R. and Snow, C. (1986), "Organisations: new concepts for new forms", *California Management Review*, 28, p. 62-73.

Milgrom, P. and Roberts, J. (1990), "The Economics of Modern Manufacturing: Technology, Strategy, and Organisation", *American Economic Review*. 80, June, p. 511-528.

Milgrom, P. and Roberts, J. (1992), *Economics, Organisation and Management,* Prentice-Hall.

Mincer, J. (1970), "The distribution of labor incomes: a survey with special reference to the human capital approach", *Journal of Economic Literature*, 8(1), p. 1-26.

Mincer, J. (1974), *Schooling Experience and Earnings*, National Bureau of Economic Research, New York.

Mitchell, D. (1991), *Income Transfers in Ten Welfare States,* Avebury.

Mitchell, D., Harding, A. and Gruen, F. (1994), "Targeting Welfare", *Economic Record*, 70 (10), p. 315-340.

Moe, T. (1984), "The New Economics of Organisation", *American Journal of Political Science*, 28(4), p. 739-777.

Mol, A., Duncan, L. and Lauber, V. (2000), "Introduction", in Mol, A., Duncan, L. and Lauber, V. (eds.), *The Voluntary Approach to Environmental Policy: Joint Environmental Policy-making in Europe*, Oxford, Oxford University Press.

Moldofsky, N. (1989), *Order – With or Without Design: selections from F. A. Hayek's Contribution to the Theory and Application of Spontaneous Order,* Centre for Research into Communist Economies.

Mooney, G.H. (1979), "Values in health care", in Lee, K. (ed.), *Economics and Health Planning*, London, Croom Helm.

Mueller, D.C. (2003), *Public Choice III*, Cambridge, Cambridge University Press.

Murray C.J.L., Salomon J.A., Mathers C. (1999), "A critical examination of summary measures of population health", GPE Discussion Paper No. 12, World Health Organisation, Geneva.

Musgrave, R.A. (1959), *The Theory of Public Finance,* New York, Mcgraw-Hill.

Myles, J. (1989), *Old Age in the Welfare State: The Political Economy of Public Pensions,* Revised Edition, University Press of Kansas, Lawrence.

Nelson, K. (2004), "Mechanisms of Poverty Alleviation: anti-poverty effects of non-means-tested benefits in five welfare states", *Journal of European Social Policy*, 14(4), p.371-390.

Nelson, R.R. and Winter, S.G. (1982), *An Evolutionary Theory of Economic Change,* Cambridge, Harvard University Press.

Niskanen, W.A. (1971), *Bureaucracy and Representative Government,* Chicago, Aldine Atherton.

Nooteboom, B. (1999), *Inter-firm Alliances. Analysis and Design*, London and New York, Routledge.

North, D.C. (1981), *Structure and Change in Economic History*, New York, Norton and Co.

North, D.C. (1990), *Institutions, Institutional Change, and Economic Performance*, Cambridge, Cambridge University Press.

Nourzad, F. and Vrieze, M. (1995), "Public Capital Formation and Productivity Growth: Some International Evidence", *Journal of Productivity Analysis*, 6, p. 283-295.

Nunamaker, T. (1985), "Using data envelopment analysis to measure the efficiency of non-profit organisations: a critical evaluation", *Managerial and Decision Economics*, 6(1), p. 50-58.

O'Mahony, M. and Stevens, P. (2004), "International Comparisons of Performance in the Provision of Public Services: outcome based measures for education", National Institute of Economic and Social Research, London.

OECD (2001), *Knowledge and Skills for Life: First Results from PISA 2000,* OECD, Paris

OECD (2002), *OECD Health Data 2002*, Organisation for Economic Cooperation and Development, Paris.

OECD (2003), *Health at a Glance: OECD Indicators 2003,* Organisation for Economic Cooperation and Development, Paris.

OECD (2004a), *Education at a Glance – OECD Indicators 2004,* Organisation for Economic Cooperation, Paris.

OECD (2004b), *Learning from Tomorrow's World – First Results from PISA 2003*, OECD, Paris.

OECD (2004c), *OECD Health Data 2004,* Organisation for Economic Cooperation and Development, Paris.

Office of Water Trading (1999), "Future Water and Sewerage Charges 2000-05: Draft Determination", London, Office of Water Trading.

Olsen, J. and Peters, B. (eds.) (1998), *Lessons from Experience*, Oslo, Scandinavian University Press.

Olson, M. (1982), *The Rise and Decline of Nations: Economic Growth, Stagflation, and Social Rigidities*, New Haven, Yale University Press.

Olson, M. (1993), "Dictatorship, Democracy and Development", *American Political Science Review*, 3(87), p. 567-576.

Osborne, D. and Gaebler, T. (1992), *Reinventing Government: how the entrepreneurial spirit is transforming the public sector*, Adison Wesley, Reading Mass.

Ouchi, W.G. (1980), "Markets, Bureaucracies, and Clans", *Administrative Science Quarterly*, 25(1), p. 129-141.

Ouchi, W.G. and Bolton, M. (1988), "The logic of joint research and development", *California Management Review*, 30, p. 9-33.

Oxley, J. (1999), "Institutional Environment and the Mechanism of Governance: the Impact of Intellectual Property Protection on the Structure of Inter-firm Alliances", *Journal of Economic Behavior and Organisation*, 38, p. 283-309.

Ozcan, Y.A., Luke, R.D. and Haksever, C. (1992), "Ownership and Organisational Performance: a comparison of technical efficiency across hospital types", *Medical Care*, 30(9), p. 781-94.

Palay, T.M. (1984), "Comparative Institutional Economics: the Governance of the Rail Freight Contract", *Journal of Legal Studies*, 13, June, p. 265-288.

Palda, P. (1997), "Fiscal Churning and Political Efficiency", *Kyklos*, 50(2), p. 189-206.

Palme, J. (1990), "Pension Rights in Welfare Capitalism: The Development of Old-Age Pensions in 18 OECD countries 1930-1985", Swedish Institute for Social Research, Stockholm.

Pareto, V. (1906), *Manual of Political Economy*, 1971 translation of 1927 edition, New York, Augustus M. Kelley.

Park, S.H. (1996), "Managing an Interorganisational Network: A Framework of the Institutional Mechanism for Network Control", *Organisation Studies*, 17(5), p. 795-824.

Park, W. (1995), "International R&D spillovers and OECD economic growth", *Economic Inquiry*, 33(4), p. 571-591.

Peacock, A. and Wiseman, J. (1961), *The Growth of Public Expenditure in the United Kingdom,* Princeton University Press.

Pedersen, A.W. (1999), "The Taming of Inequality in Retirement. A Comparative Study of Pension Policy Outcomes", Fafo-report 317, Oslo.

Pedraja-Chaparro, F., Salinas-Jiménez, J. and Smith, P. (1997), "On the role of weight restrictions in data envelopment analysis". *Journal of Productivity Analysis*, 8, p. 215-230.

Penrose, E.T. (1959), *The Theory of the Growth of the Firm,* Blackwell, Oxford.

Pettigrew, A., Ferlie, E. and Mckee, L. (1992), *Shaping Strategic Change*, London, Sage.

Pfeffer, J. and Nowak, P. (1976), "Joint Ventures and Interorganisational Interdependence", *Administration Science Quarterly*, 21, p. 398-418.

Pierre, J. (2000a), *Debating Governance: Authority, Steering and Democracy*, Oxford, Oxford University Press.

Pierre, J. (2000b), "Introduction: Understanding Governance", in *Debating Governance: Authority, Steering, and Governance*, Pierre, J. (ed.), Oxford University Press, Oxford.

Piore, M.J. (1992). “Fragments of a cognitive theory of technological change and organisational structure”, in Nohria, N. and Eccles, R.G. (eds.), *Networks and Organisations: Structure, Form and Action. Boston*, MA, Harvard Business School Press, p. 430-44.

Pisano, G.P. (1990), “The R&D Boundaries of the Firm: An Empirical Analysis”, *Administrative Science Quarterly*, 35(1), p. 153-176.

Podolny, J. and Page, K. (1998), “Network Forms of Organisations”, *Annual Review of Sociology*, 24, p. 57-76.

Pollitt, C. (2003), “Decentralization: A Central Concept in Contemporary Public Management”, paper presented for the European Group of Public Administration Conference, Lisbon.

Pollitt, C. and Bouckaert, G. (2000), *Public Management Reform*, Oxford, Oxford University Press.

Pollitt, C. and Bouckaert, G. (2004), *Public management reform: a comparative analysis*, 2nd edition, Oxford, Oxford University Press.

Pollitt, C., Birchall, J. and Putman, K. (1998), *Decentralising Public Service Management*, London, Macmillan.

Porter, M. and Fuller, M. (1986), “Coalitions and global strategy”*,* in Porter, M., *Competition in Global Industries,* Boston, Harvard Business School.

Powell, W.W. (1987), “Hybrid Organisational Arrangements: new forms or transitional development”*, California Management Review*, 30(1), p. 67-87.

Powell, W.W. (1990), “Neither Market nor Hierarchy: Network Forms of Organisation”, in Cummpings, L.L. and Staw, B. (eds), *Readings in Organisational Behavior*, Greenwich, JAI Press, 12, p. 295-336.

Powell, W.W. (1996), “Inter-Organisational Collaboration in the Biotechnology Industry”, *Journal of Institutional and Theoretical Economics*, 152(1), p. 197-215.

Raynaud, E. (1997), *Propriété et exploitation partagée d’une marque commerciale: aléas contractuels et ordre privé*, Ph.D. Université de Paris (Panthéon-Sorbonne).

Reifschneider, D. and Stevenson, R. (1991), ”Systematic departures from the frontier - a framework for the analysis of firm inefficiency”, *International Economic Review,* 32(3), p. 715-724.

Rhodes, R.A.W. (1997), *Understanding Governance: Policy Networks, Governance, Reflexivity and Accountability*, Buckingam, Open University Press.

Richardson, G.B. (1972), “The Organisation of Industry”, *The Economic Journal*, September, p. 883-896.

Richardson, G.B. (1998), "Some Principles of Economic Organisation", chapter 3 in Foss, N.J. and Loasby, B.J (eds.), *Economic Organisation, Capabilities and Co-Ordination,* London, Routledge, p. 44-62.

Robson, M. (1993), "Federal funding and the level of private expenditure on basic research", *Southern Economic Journal*, 60(1), p. 63-71.

Romer, P. (1990), "Endogenous technological change", *Journal of Political Economy*, 99(5), p. S71-S102.

Romero de Avíla, D. and Strauch, R. (2003), "Public Finances and Long-term Growth in Europe – Evidence from a Panel Data Analysis", ECB Working Paper n. 246.

Rosenau, J.N. (1992), "Governance, Order, and Change in World Politics", in Rosenau, J.N. and Czempiel, E-O. (eds.), *Governance without Government: Order and Change in World Politics*, Cambridge, Cambridge University Press.

Rosenau, J.N. (2000), "Change, complexity and governance in globalizing space", in Pierre, J. (ed.), *Debating governance: authority, steering and democracy*, Oxford, Oxford University Press.

Rosenau, J.N. and Czempiel, E-O (eds.) (1992), *Governance without government: Order and change in world politics*, Cambridge, Cambridge University Press.

Rubin, P.H. (1978), "The Theory of the Firm and the Structure of the Franchise Contract", *Journal of Law and Economics*, 21(1), p. 223-233.

Ryan, M. and Farrar, S. (2000), "Using conjoint analysis to elicit preferences for health care", *British Medical Journal*, 320, p. 1530-1533.

Sako, M. (1998), "Does Trust Improve Business Performance?", in Lane, C. and Backmann, R. (eds.), *Trust In And Between Organisations,* Oxford, Oxford University Press.

Samuelson, P. (1954), "The pure theory of public expenditure", *Review of Economics and Statistics*, 36(4), p. 387-389.

Sanchez-Robles, B. (1998), "Infrastructure investment and growth: some empirical evidence", *Contemporary Economic Policy*, 16, p. 98-108.

Sauvée, L. (2002), "Governance in Strategic Networks", WP, ISAB, 19 pages.

Savas, E.S. (2000), *Privatization and public-private partnerships*, New York, Chatham House.

Shleifer, A. (1985), "A Theory of Yardstick Competition", The RAND Journal of Economics, 16(3), p. 319-327.

Schmidt, P. (1985), "Frontier production functions", *Econometric Review*, 4, p. 289-328.

Schmidt, P. and Lin, T. (1984), "Simple tests of alternative specifications in stochastic frontier models", *Journal of Econometrics*, 24, p. 349-361.

Schneider, V. and Kenis, P. (1996), "Verteilte Kontrolle: Institutionelle Steuerung in modernen Gesellschaften", in Kenis, P. and Schneider, V. (eds.), *Organisation und Netzwerk: Institutionelle Steuerung in Wirtschaft und Politik*, Frankfurt, Campus.

Schultz, T.W. (1961), "Investment in human capital", *American Economic Review*, 51(1), p. 1-17.

Schultz, T.W. (1963), *The Economic Value of Education*, New York, Culombia University Press.

Self, P. (1993), *The Politics of Public Choice*, Basingstoke, Macmillan.

Sen, A. (1999), *Development as Freedom*, New York: Alfred A. Knopf.

Sexton, T. R. (1986), *Measuring Efficiency: An Assessment of Data Envlelopment Analysis*, San Francisco, Jossey Bass, Inc.

Shaw, R., Dolan, P., Tsuchiya, A., Williams, A., Smith, P. and Burrows, R. (2001), "Development of a questionnaire to elicit public preferences regarding health inequalities", *Occasional Paper*, Centre for Health Economics, University of York.

Shephard, R. W. (1970), *Theory of Cost and Production Functions*, Princeton University Press.

Shioji, E. (2001), "Public Capital and Economic Growth: A Convergence Approach", *Journal of Economic Growth*, 6, p. 205-227.

Silverman, E. and Skinner, J. (2000), "Are for-profit hospitals really different? Medicare upcoding and market structure", mimeo, Hanover, Dartmouth College.

Simar, L. (1992), "Estimating efficiencies from frontier models with panel data: a comparison of parametric, nonparametric and semi-parametric methods with bootstrapping", *Journal of Productivity Analysis*, 3, p. 171-203.

Simar, L. and Wilson, P.W. (2000). "A general methodology for bootstrapping in non-parametric frontier models," *Journal of Applied Statistics*, 27(6), p. 779-802.

Simar, L. and Wilson, P.W. (2002), "Estimation and Inference in Two-stage Semi-parametric Models of Production Processes", Louvain, Institut de Statistique, Universite Catholique de Louvain.

Simon, H.A. (1951), "A Formal Theory of the Employment Relationship", *Econometrica*, 19 (3), p. 293–305.

Simon, H.A. (1991), "Organisations and Markets", *Journal of Economic Perspectives*. 5(2), p. 25-44.

Smith, A. (1776), *An Inquiry into the Nature and Causes of the Wealth of Nations*, London: Methuen, Edwin Cannan (ed.), (1904), Fifth edition.

Smith, P**.** and Street, A. (2005), "Measuring the efficiency of public services: the limits of analysis", *Journal of the Royal Statistical Society,* Series A, 168, p. 401-417.

Smith, P. and Street, A. (2006), "Report on Workshop on Measurement of Non-market Output and Education", Workshop jointly organised by the OECD, The United Kingdom Office for National Statistics and the Government of Norway.

Spicker, P. (2000), *The Welfare State: a General Theory,* London, Sage.

St. Aubyn, M. (2002), "Evaluating Efficiency in the Portuguese Health and Education Sectors", paper presented to the conference "Desenvolvimento Económico Português no Espaço Europeu: Determinantes e Políticas", Banco de Portugal, May, Fundação Calouste Gulbenkian, Lisbon.

Stebbing, L.S. (1933), *A Modern Introduction to Logic,* Methuen, London.

Stevens, P. (2005a), "A stochastic frontier analysis of English and Welsh universities", *Education Economics*, 13, p. 355-374.

Stevens, P. (2005b), "Assessing the performance of local government", *National Institute Economic Review*, p. 90-101.

Stiglitz, J.E. (2000), *Economics of the Public Sector,* W.W. Norton and Company, 3rd edition.

Stinchcombe, A. (1985), "Contracts as hierarchical documents", in Stinchcombe, A. and Heimer, C. (eds.), *Organisation Theory and Project Management,* Oslo, Norwegian University Press.

Streeck, W. and Schmitter, P.C. (1985), "Community, Market, State – and Associations? The Prospective Contribution of Interest Governance to Social Order", in Streeck W. and Schmitter P.C., *Private Interest Government: Beyond Market and State*, Sage, London.

Talbot, C. (2004), "Executive agencies: have they improved management in government?", *Public Money and Management*, 24(2), p. 104-112

Tanzi, V. and Chalk, N. (2000), "Impact of Large Public Debt on Growth in the EU: A Discussion of Potential Channels", in *European Economy*, 2, p. 23–43.

Tanzi, V. and Schuknecht, L. (1997a), "Reconsidering the Fiscal Role of Government: The International Perspective", *American Economic Review*, 87(2), p. 164-168.

Tanzi, V. and Schuknecht, L. (1997b), *Reforming Government: An overview of recent experience,* European Journal of Political Economy, 13, p. 395-417.

Tanzi, V. and Schuknecht, L. (2000), *Public spending in the 20th century: a global perspective,* Cambridge University Press.

Tanzi, V. and Schuknecht, L. (2003), "Public Finances and Economic Growth in European Countries", in "Fostering Economic Growth in Europe", conference volume of the 31st Economics Conference of the Oestereichische Nationalbank, Vienna, 2003, 178-196.

Tanzi, V. and Zee, H. (2000), "Taxation and the Household Saving Rate: Evidence from OECD Countries", *Banca Nazionale del Lavoro*, Quarterly Review, 212.

Teece, D. and Pisano, G.P. (1994), "The Dynamic Capabilities of Firms: An Introduction", *Journal of Economic Behavior and Organisation*, 27(3), p. 537-556.

Teixeira, R.A. (1992), *The Disappearing American Voter*, Brookings, Washington, DC.

Tenbensel, T. (2005), "Multiple Modes of Governance: Disentangling the alternatives to markets and hierarchies", *Public Management Review*, 7(2), p. 267-288.

Thanassoulis, E. (2001), "Introduction to the Theory and Application of Data Envelopment Analysis", Dordrecht, Kluwer.

Thompson, G., Frances, J., Levacic, R. and Mitchell, J. (1991), *Markets, Hierarchies and Networks: the coordination of social life,* London, Sage.

Thöne, M. (2004), "Wachstums- und nachhaltigkeitswirksame öffentliche Ausgaben ("WNA")", Study commissioned by the German Federal Ministry of Finance, final version, FiFo Köln (Cologne Center for Public Finance).

Thorelli, H.B. (1986), "Networks: Between Markets and Hierarchies", *Strategic Management Journal*, 7(1), p. 37-51.

Tirole, J. (1986), "Hierarchies and Bureaucracies: On the Role of Collusion in Organisations", *Journal of Law, Economics, and Organisations*. 2(2), p. 181-214.

Tirole, J. (1988), *The Theory of Industrial Organisation*, Cambridge MA, MIT Press.

Titmuss, R. (1958), *Essays on the Welfare State,* Allen & Unwin, London.

Titmuss, R. (1974), *Social Policy,* Allen & Unwin, London.

Toivanen, O. and Niininen, P. (1998), "Investment, R&D, subsidies, and credit constraints", Working Paper, Department of Economics MIT and Helsinki School of Economics.

Treib, O., Bähr, H. and Falkner, G. (2004), "Modes of Governance: Old and New: A Note Towards Conceptual Clarification", paper presented at the kick-off workshop of

NEWGOV cluster one: “Emergence, Evolution and Evaluation“, Brussels, December 10.

Treib, O., Bähr, H. and Falkner, G. (2005), “Modes of Governance: A Note Towards Conceptual Clarification”, European Governance Papers No. N-05-02.

Udehn, L. (1996), *The Limits of Public Choice*, London, Routledge.

Van Rijckeghem, C. and Weder, B. (2002), “Bureaucratic Corruption and the Rate of Temptation: Do Wages in the Civil Service Affect Corruption and by How Much?”, in G. Abed and S. Gupta (eds), *Governance, Corruption and Economic Performance*, IMF: Washington DC, p. 59-88.

Vanden Eeckaut, P., Tulken, H., and M-A Jamar (1993), “Cost Efficiency in Belgian Municipalities,” in Fried, H.O., Lovell, C.A.K., and Schmidt, S.S. (eds), *The Measurement of Productive Efficiency: Techniques and Application*, New York, Oxford University Press.

Vanhoudt, P.; Mathä, T. and Smid, B. (2000), “How productive are capital investments in Europe?”, EIB Papers, 5(2), p. 81-106.

Von Hagen, J., Hallerberg, M. and Strauch, R. (2004), “The Design of Fiscal Rules and Forms of Governance in European Union Countries”, mimeo.

Walker, N. (2005), *The Legal Dimension of New Modes of Governance: Background Information*, paper presented at NEWGOV Consortium Conference, 30-31 May 2005, Florence.

Wallsten, S. (1999), “Do government-industry R&D programs increase private R&D?: The Case of the Small Business Innovation Research Program”, Department of Economics Working Paper, Stanford University, Stanford CA.

Wellman, B. (1983), “Network Analysis: Some Basic Principles”, in Collins, R. (ed.), *Sociological Theory*, San Francisco, Jossey-Bass, p. 155-200.

Whinston, M. (2003), “On the Transaction Costs Determinants of Vertical Integration”, *Journal of Law, Economics and Organisation*, 19(1), p. 1-23.

White, H.C. (1981), “Where Do Markets Come From?”, *American Journal of Sociology*, 87, November, p. 548-577.

WHO (2000), *The World Health Report – Health systems: improving performance,* World Health Organisation, Geneva.

WHO (2003), *The world health report 2003 – shaping the future,* The World Health Organisation, Geneva.

Williamson, O.E. (1967), “Hierarchical Control and Optimum Firm Size”, *Journal of Political Economy*, 75(2), p. 123-138.

Williamson, O.E. (1971), "The Vertical Integration of Production: Market Failure Considerations", *American Economic Review*, 61, p. 112-123.

Williamson, O.E. (1975), *Markets and Hierarchies: Analysis and Anti-Trust Implications,* New York: Free Press.

Williamson, O.E. (1985), *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting,* New York, The Free Press-McMillan.

Williamson, O.E. (1991), "Comparative Economic Organisation: The Analysis of Discrete Structural Alternatives", *Administrative Science Quarterly*, 36(2), p. 269-296.

Williamson, O.E. (1996), *The Mechanisms of Governance*, Oxford, Oxford University Press.

Williamson, O.E. (2004), "The Economics of Governance", unpublished paper available at http://groups.haas.berkeley.edu/imio/owilliam101404.pdf.

Windhoff-Heretier, A. (1987), "Emergence and Evolution of Modes of Governance: Designing a Treaty-based Framework – Towards an Index of Integration", paper presented at NEWGOV Cluster 1 Meeting, May 30/31, Florence.

Wolf, M. (1988), *Markets or Government: Choosing Between Imperfect Alternatives*, Cambridge, MIT Press.

Wolfe, B.L. and Zuvekas, S. (1995), Nonmarket outcomes of schooling", Discussion Paper 1065-95, Institute for Research on Poverty, University of Wisconsin, Madison, WI).

Yarbrough, B. and Yarbrough, R. (1988), "The Transaction Structure of the Firm: A Comparative Survey", *Journal of Economic Behavior and Organisation*, 10(1), p. 1-28.

Yi, K. and Kocherlakota, N. (1996), "Is there endogenous long-run growth? Evidence from the U.S. and the U.K.", Federal Reserve Bank of New York, 17(12).

Zagler, M. and Durnecker, G. (2003), "Fiscal Policy and economic growth", *Journal of Economic Surveys*, 17(3), p. 397-418.

Zifack, S. (1994), *New Managerialism: Administrative Reform in Whitehall and Camberra*, Buckingham, Open University Press.